2017

Northumbria University

Luca Miorelli

THE DEVELOPMENT OF MORPHO-SYNTACTIC COMPETENCE IN ITALIAN-SPEAKING CHILDREN: A USAGE-BASED APPROACH

Volume I of II

Supervisors:
Prof. Ewa Dąbrowska
Dr James Street

The Development of Morpho-Syntactic Competence in Italian-Speaking Children: A Usage-Based Approach

Luca Miorelli

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle upon Tyne for the degree of Doctor of Philosophy

Research undertaken in the Faculty of Arts, Design & Social Sciences

October 2017

ABSTRACT

Usage-Based scholars (e.g. Lieven et al., 2009) have shown that children's early grammar is characterisable as knowledge of lexically-specific patterns (***kick** KICKEE*) learnt from previously encountered strings (***kick** it*). Experimental research (e.g. Lewis, 2009) has shown that children younger than four years cannot use nonce verbs in constructions in which they have never experienced them. Productivity with nonce verbs slowly improves throughout the preschool years, as adultlike schemas (e.g. *AGENT-PROCESS-PATIENT*) gradually emerge in ontogeny (Tomasello, 2006b).

However, such results are overwhelmingly based on studies of English-speaking children and it is unclear how well they generalise to other languages. The research presented in this thesis enquired into whether a Usage-Based Approach could account for the acquisition of Italian.

A longitudinal study investigated whether the spontaneous production of an Italian-speaking two-year-old could be accounted for in terms of lexically-specific units instantiated in the concrete strings he had previously experienced. An experimental study tapped into the development of 2;02-to-5;0-year-old Italian-speakers' productivity with past participles and the transitive construction using both a nonce verb and a familiar verb.

Results on syntactic development were consistent with previous findings regarding English-speaking children (Akhtar, 1999; Lieven et al., 2009). The overwhelming majority of the child's spontaneous production (82%) could be derived from previously encountered lexically-specific patterns. In the experimental setting, children younger than four years could not produce adultlike transitive sentences with a nonce verb they had not experienced in that construction.

As for morphological productivity, even two-year-olds used the nonce verb productively. Such results are discussed in terms of how the co-occurrence of high type and token frequency that characterises the Italian morphology may facilitate form-function mapping.

Overall results are consistent with Usage-Based Models, suggesting that such approaches have cross-linguistic validity.

Page intentionally left blank

TABLE OF CONTENTS

VOLUME I

Part II:

The Spontaneous Productions of a Two-year-old Child

Part IV:
Conclusions

VOLUME II

Part V:

Appendix

Page intentionally left blank

LIST OF ABBREVIATIONS (A-H)

1st.conj	Conjugation I
2nd.conj	Conjugation II
3rd.conj	Conjugation III
AA	Anaphoric Ante-position
ACC	Accusative
ADJ	Adjective
ADV	Adverb
APA	Adaptive Processing Approach
ASC	Argument Structure Construction
AUX	Auxiliary Verb
CART(s)	Classification and Regression Tree(s)
CDS	Child Directed Speech
CG	Cognitive Grammar
CI(s)	Confidence Interval(s)
CL	Cognitive Linguistics
clit.	Clitic
CM	Competition Model
cmc	Caused-Motion Construction
CMH	Critical Mass Hypothesis
COND	Conditional
conj.I	Conjugation I
conj.II	Conjugation II
conj.III	Conjugation III
COP	Copula
DAT	Dative
DIM	Diminutive
DO	Direct Object
DV	Dependent Variable
E.	Experimenter
ecp	Early Competence Point
ENDR	Endearment
F	Feminine
fem	Feminine
FL	Faculty of Language
FLSS	Fully Lexically-specific String(s)
FOC	Focalisation
FP	Full Productivity
GG	Generative Grammar
hcf	Hard Constructional Fail(s)
HT	Hanging Topic

LIST OF ABBREVIATIONS (I-S)

IMP	Imperative
IMPERF	Imperfect
IMPRS	Impersonal
IND	Indicative
IO	Indirect Object
IV	Independent Variable
KA	knowledge Access
LA	Language Acquisition
LD	Left Dislocation
LOC	Locative
LSP	Low Scope Productivity
M	Masculine
masc	Masculine
MP	Minimalist Program
N	Noun
NML	Nominal
NOM	Nominative
NP	Noun Phrase
o	(as in oVS): Accusative Clitic Pronoun
O	(as in SVO): object-NP
opc	Object Permanence Concept
PDP	Parallel Distributed Processing
PL	Plural
PP	Prepositional Phrase
PRON	Pronoun
PRS	Present Tense
PTCP	Past Participle
REFL	Reflexive
S	Subject
SBJ	Subject
SBJV	Subjunctive
scf	Soft Constructional Fail(s)
SG	Singular
SS	Sanctioning Structure
SwS	Schema with Slot

LIST OF ABBREVIATIONS (T-Z)

ToM	Theory of Mind
TS	Target Structure
TV	Thematic Vowel
tvl	Test di Valutazione del Linguaggio (Test for the Assessment of Language Development)
UB	Usage-Based
UBA	Usage Based Approach
UE	Usage Event
V	Verb
VHI	Verb Island Hypothesis
VP	Verb Phrase
WO	Word Order

Page intentionally left blank

Acknowledgements

The following research has been undertaken between October 2013 and October 2017 at the Northumbria University of Newcastle upon Tyne (UK), in the Faculty of Arts, Design and Social Sciences. The research was funded by a Northumbria Research Scholarship for the period October 2013-October 2016 and self-funded during the period October 2016-October 2017.

My gratitude goes to my supervisors Prof. Ewa Dąbrowska and Dr James Street, who have provided me with their (wise and precious) guidance and feedback throughout these years.

This research would have not been possible without the help and collaboration of the institutions, nurseries, teachers, carers, parents and children who participated in, collaborated with and authorised the study. Although I cannot name any of them for privacy reasons, my gratitude to them is immense, as without such amazing volunteers it would not be possible to conduct scientific research.

A final thanks goes to Matthew Mulhall for the precious discussions on my work, the Italian language, and whatever else we talked about and to my family, Eve and all my friends.

The final thought goes to the memory of Filippo "Phil" Martello (1986-2011).

Page intentionally left blank

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others. The research was conducted under the supervision of Prof. Ewa Dąbrowska and Dr James Street.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the Faculty Ethics Committee (22.11.2014, Longitudinal Study and 03.06.2014, Experimental Study).

I declare that the Word Count of this Thesis is 84,591 words, excluding table of contents, list of abbreviations, appendix, footnotes, references, tables and captions of figures and tables.

Name: *Luca* Miorelli

Signature:

Date: *4th of October, 2017*

Page intentionally left blank

A Filippo "Phil" Martello

Page intentionally left blank

The Development of Morpho-Syntactic Competence in Italian-Speaking Children: A Usage-Based Approach

Volume I

Page intentionally left blank

0.

INTRODUCTION

The research that is to be presented in the current thesis investigates the acquisition of Italian as a first language. This introductory chapter walks the reader through a brief discussion as to why studying language acquisition is relevant to a general understanding of what language is and also contextualises the current research within the broader field of linguistics and the so-called *nature-vs.-nurture* debate that characterises it. A brief illustration of the structure of the thesis ends the current chapter.

0.1.WHY STUDY LANGUAGE ACQUISITION

Language is a distinguishing feature of human beings. We are exposed to language during virtually every moment of our life, from social interactions to more solitary activities (watching television, reading, etc.).

We produce and understand many novel utterances on a daily basis and we are able to do this because our language use follows certain conventions. Indeed, we are able to do so because we master a complex system of phonological, semantic, morphological and syntactic patterns (we have a grammar of our native language). As if this were not enough, we also process language very quickly. In an ordinary conversation, English-speakers utter an average of 210 words per minute (Tauroza & Allison, 1990[1]). Yet, possibly none of us has ever been explicitly taught his or her own mother tongue. Children develop grammatical competence from the finite number of sentences they encounter (and use) during development (Dąbrowska, 2014).

Language is thus a (complex) mental and cognitive system crucial to human beings (Goldberg, 2006). Grammar can be thought of as the way in which we represent such a mental system (Croft & Cruise, 2004).

In order to develop a psychologically realistic theory of what such a mental and cognitive system is, researchers must account for the ways in which language manifests:

[1] Mentioned in Dąbrowska (2014).

a) Productivity:

> "We produce and understand an indefinite number of novel utterances"
>
> (Dąbrowska, 2014, p. 617).

b) Learnability: children learn the (ambient) languages that are spoken around them (Croft & Cruise, 2004; Dąbrowska, 2014).

c) Real-time processing: we process language very quickly and under conditions that are far from ideal (when we are in a noisy pub, for example; see Dąbrowska, 2014).

d) Language change: speakers' grammatical knowledge changes over time (Croft & Cruise, 2004). Such changes manifest on two levels: on the personal level (during a single person's life span words acquire new meanings and partially lose old ones) and on a social and historical level (languages change during the course of centuries).

The study of Language Acquisition (LA, henceforth) contributes to the understanding of what linguistic knowledge is in three ways:

i). It straightforwardly provides an insight into point (b), as the study of LA is about how human beings learn languages.

ii). In order to learn a language, children must make sense of the utterances they hear, even those they are hearing for the first time. Furthermore, eventually children will have to develop the ability to produce novel utterances they have never encountered. Hence, studying LA sheds light on language productivity (point a).

iii). Finally, children's linguistic improvement over time is supposedly related to qualitative changes in their mental representation of their language. Investigating LA provides insights into language ontogeny (point d).

0.2. NATURE vs. NURTURE

Virtually every branch of Linguistics is characterised by the *nativists versus non-nativists debate*, and LA is no exception (O'Grady, 2012). **Nativist researchers** claim that language is innate (i.e. part of human beings' genetic endowment) and its acquisition is biologically programmed in a similar way to puberty and psychomotor development. **Usage-based (UB,** henceforth**) scholars,** whose research posits itself within *non-nativist* theories, question the putative innate and biologically programmed nature of language and LA. Instead, they claim that languages are learnt and used by applying general cognitive skills (such as generalisation, intention reading and cultural learning) to the task of communicating.

0.2.1. THE USAGE-BASED APPROACH

Usage-based approaches (UBAs, henceforth) are rooted in Construction (Goldberg, 1995, 2006); and Cognitive (Langacker, 1987, 1991, 2000, 2008, 2010) Grammar. According to these approaches, speakers' knowledge of their language is describable as mastery of a network of highly interconnected constructions. Constructions are form-meaning pairings of different sizes (from single words to whole sentences) and different degrees of abstraction (from fully lexically-specific to fully schematic; table 0.1). Words, semi-productive patterns and grammatical (regular) patterns are all form-meaning pairings lying on a continuum of complexity, abstraction and generalisability, whose acquisition is underpinned by the same processes of form-meaning mapping (lexicon-syntax continuum; see Langacker, 2008; Dąbrowska, 2004).

Table 0.1: Constructions, form-meaning pairings. Elements in small letters are fully lexically-specific concrete words, whereas CAPITALS indicate SLOTS (that is, more schematic, lexically-unspecified semantic/functional generalisations).

	SYMBOLIC UNITS		
	simple	complex	
fully concrete	banana	kicked	I make those
partially schematic	b..........	PROCESS-ed	MAKER make THING
fully schematic	THING	PROCESS-INFLECTION	AGENT-PROCESS-PATIENT

Fully abstract (or fully-schematic) complex constructions (such as the *AGENT-PROCESS-PATIENT* construction) correspond to what are traditionally regarded as morpho-syntactic rules (Langacker, 2008) and can be thought of as templates that link a form (the sequence NP1 V NP2) to a meaning (NP1 acts upon NP2). These units are called constructional schemas and are the by-products of generalisations that speakers draw from the concrete language they encounter (Langacker, 2000, 2008; Croft & Cruise; 2004).

Children initially learn chunks of language of any size, from single words to whole sentences. These chunks are form-meaning pairings (phonological shapes are mapped onto meanings), whose internal structure may initially be "parsed" at different degrees of detail (*givemethat* but *where's* < *mum* and *I* < *want* < *pizza*). Once children have analysed at least part of their formulaic strings, around the age of two, they develop lexically-bound schemas (Tomasello, 2003). The latter are constructions that are bound to specific lexical items (*WANTER-**want**-THING*) and are learnt by drawing analogies between concrete strings (*I want pizza*; *we want this*) that instantiate them, on the basis of lexical (/wɒnt/), functional (obtaining something or communicating desire for something), semantic (about wanting something) and distributional (the *WANTER* takes pre-verbal position) similarities. Once children have learnt many lexically-bound schemas (*WANTER-want-THING*; *EATER-eat-THING*), they draw analogies across them and develop constructional schemas (*AGENT-PROCESS-PATIENT*), whose acquisition yields a more schematic and adultlike language.

0.2.2. THE NATIVIST APPROACH

The hypothesis that language knowledge is biologically endowed originates in Chomsky's (1959, 1965, 1975, 1980) early works. The human mind, he claims, is composed of modules, one of which is the Faculty of Language (FL, henceforth). The FL is a

> "language organ, in the sense in which scientists speak of the visual system, or immune system, or circulatory system, as organs of the body"
>
> (Chomsky, 2000, p. 4).

Human beings are not endowed with the grammar of a specific language (English, Norwegian, etc.), but with Universal Grammar (UG), a characterisation of the kind of knowledge that constitutes the FL (Chomsky, 1986). UG is made of principles, which are *highly abstract* and language-specific properties shared by all languages.

For instance, the X-bar theory is the principle which determines that every lexical category X (say a verb, V) heads a projection X' (X-bar = V-bar) made of X and its complement (V + NP-object). X' can then further project into a maximal projection X'', made of X' and a specifier (Spec.) (Chomsky, 1986). The X-bar theory establishes a hierarchical structure within phrases (PP, VP, etc.), which is traditionally represented as in fig. 0.1.

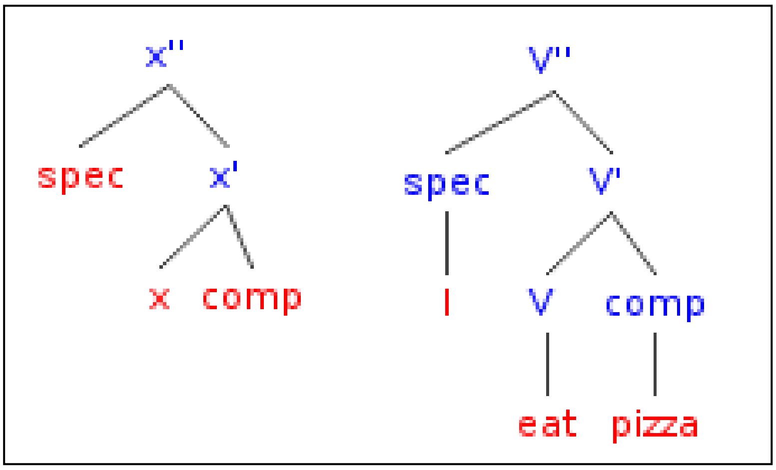


Figure 0.1: the x-bar structure.

Within some principles, UG allows a certain parametric variation. Parameters are binary options (+/-), which are internally associated with a given principle (Chomsky, 1986; Haegeman, 1994). Simply put, there is only one underlying human grammar based on universal principles. Typological variation is the by-product of the parametric values each language assigns to principles and the specific lexical features of that language (see Chomsky, 2000). The parameter that yields cross-linguistic variation within X-bar Theory is the *head-parameter*. Such a parameter determines the position of the head with respect to its complement(s) and can take either *head-first* or *head-last* value. The former setting yields VO languages (e.g. English), the latter yields OV languages (e.g. German).

Children have in-built knowledge of principles and their parametric variation. LA is about setting parameters during a developmental path that is guided and constrained by UG (Chomsky, 1986, 2000; Yang & Roeper, 2011). Once a parameter is set, all aspects of language governed by that parameter are acquired instantly (Chomsky, 1986; Radford, 1990). If an English-speaking child hears *eat soup*, s/he notices that the order VO is used; the head parameter is then set as head-first. Such a setting is then automatically applied to all lexical heads: the child now knows that, for example, PPs have the order P-NP, without needing to be exposed to instances of PPs (Chomsky, 1986).

However, two main methodological and theoretical issues haunt generativist research. Firstly, it is still unclear how many parameters there are and what they may be (see Dąbrowska 2015; Tomasello 2005). Secondly, throughout time, UG principles have undergone substantial revisions (Dąbrowska, 2015), up to the point where they appear to be dismissed (or at least greatly reduced) by Chomsky's (1995) Minimalist Program (MP, henceforth).

Chomsky (1995, p. 30) claims that

> "language consists of a lexicon and a computational system"

and that the latter generates expressions out of the former.

Indulging in some degree of approximation, Hauser, Chomsky and Fitch (2002) distinguish between two parts that make up the FL. The narrow FL is the computational system that underpins language and its knowledge and is unique to humans. The broad FL underpins those aspects of language which are not language-specific and possibly shared by other animals. According to Chomsky (Chomsky, 1995, 2000; Hauser et al. 2002) the narrow FL solely consists of mechanisms of recursion, which manifests through the operations of Merge and Move (see Chomsky, 1995 for details).

Thus, the MP appears to greatly reduce the type and amount of innate **grammatical** knowledge, sometimes up to the point where some scholars (e.g. Longa & Lorenzo, 2008) suggest that the nature and base of the FL no longer need be language-specific;

> […] the nature of the FL is not different from that of the external systems. It continues to be a universal and, conceivably, innate base, but nevertheless not a specifically grammatical one.
>
> (Longa & Lorenzo, 2008, p. 546).

For instance, Hauser et al. (2002) claim that recursion is exploited for numerical/quantificational operations as well as in language.

Thus, once such an approach is adopted, the extent to which one can still speak of innateness, or of the modularity of language, is at best controversial and, more importantly, it is not clear the role innate structures have in acquisition (O'Grady, 2012). Chomsky (1995, 2000, 2004, 2005, 2007) seems to imply that knowledge of permitted parameters is what remains as innate linguistic knowledge: even under the MP, LA is still

> "a matter of parameter setting"
>
> (Chomsky, 2000, p. 8)

However, LA cannot be about parameter setting only, as languages have idiosyncratic constructions (*the Xer the Yer*), idioms (*X Kicked the bucket)*, lexical items and irregular morphology (*throw, threw, thrown*) that cannot be innate. Indeed, such linguistic phenomena are often so idiosyncratic and language-specific that it is difficult to see how they could be specified in a biologically endowed LF shared by all humans. For instance, different languages form the past participle of verbs in different ways (*-to* in Italian, -/d/, –/t/ and -/ɪd/ in English). To posit that such specific rules of past participle formation are part of children's genetic endowment would be unrealistic (see Yang & Roeper, 2011). Chomsky (1965) distinguishes between core language (UG) and periphery (lexicon, semi-productive patterns, patterns of morphological inflections, etc.). Syntax is part of UG, lexicon is not (the so-called *lexicon-syntax discontinuity*). Those aspects that fall within the periphery are learnt by means of general cognitive abilities. For instance, in order to learn how to form past participles (part of the periphery in generativist terms) English-speaking children must generalise from specific verbs (*kicked, picked*) in order to infer the pattern *PROCESSed* (see Yang & Roeper, 2011).

0.2.3. POVERTY OF THE STIMULUS?

The stronghold of arguments for innateness is the so-called **poverty of the stimulus**[2]. Grammatical knowledge, Chomsky (1959, 1965) suggests, is a system of underlying rules, which is too complex to be learnt from the limited and often degraded input to which children are exposed. The strongest argument

[2] For a thorough, yet fairly succinct review of "poverty of the stimulus" issues raised by nativist theories, refer to Shwartz and Sprouse (2013).

for the insufficiency of the input is the fact that speakers rely on structure-dependent operations - that apply to syntactic categories such as phrases (NP, VP, PP) – rather than on structure-independent operations – based on principles such as the specific position of words. The former are not acquirable through general, non language-specific learning (see Chomsky, 1959, 1972, 1980; Aitchison, 2011). Chomsky's (1975, pp. 30-32) most used example goes as follows:

Suppose that a child heard (1) and (2), s/he could assume that questions are formed by fronting the first verb (a structure-independent operation).

1. The boy is angry.
2. Is the boy _ angry?

However, such a hypothesis would derive the ungrammatical (4) from (3).

3. The boy who is screaming is angry.
4. *Is the boy who _ screaming is angry?

To deduce the right rule, the child must infer that it is the verb of the main clause that moves. This implies that the child is able to parse (3) into abstract phrases (5) and front the verb of the main clause (6) (Chomsky, 1975, p. 32).

5. [The boy who is screaming] [is] [angry].
6. Is the boy who is screaming _ angry?

Yet, children never seem to produce (4)-like sentences, which suggests that they know that only structure-dependent operations are possible (Chomsky, 1975; Thomas, 1993). Nativists claim that children do not need to learn that operations are structure-dependent, because such knowledge is part of their genetic endowment:

> "languages are learnable because there is little to learn"
>
> (Chomsky, 2000, p. 124).

Such claims run into two main problems. Firstly, Ambridge, Rowland and Pine (2008) showed that children do indeed make those kinds of errors (**can the boy who _ run fast can jump high?*). Secondly, the poverty of the stimulus argument stands only if one assumes that questions like (6) are formed by fronting the verb of the main clause from an original structure similar to the one in (5). However, if a UBA is taken, the poverty of the stimulus hypothesis has no reason to exist. Questions are not formed by fronting the (auxiliary) verb of the main clause through structure-dependent operations. Instead, questions are constructions (form-meaning pairings) which are learnt in the same way as other constructions: by drawing generalisations from the concrete strings that instantiate them (Goldberg, 2006; Dąbrowska, 2015). As fig. 0.2 shows, children do not need to analyse sentences into abstract phrases, but merely establish form-function correspondences on the basis of which they can develop more abstract templates and gradually acquire competence of the structure-dependent nature of language.

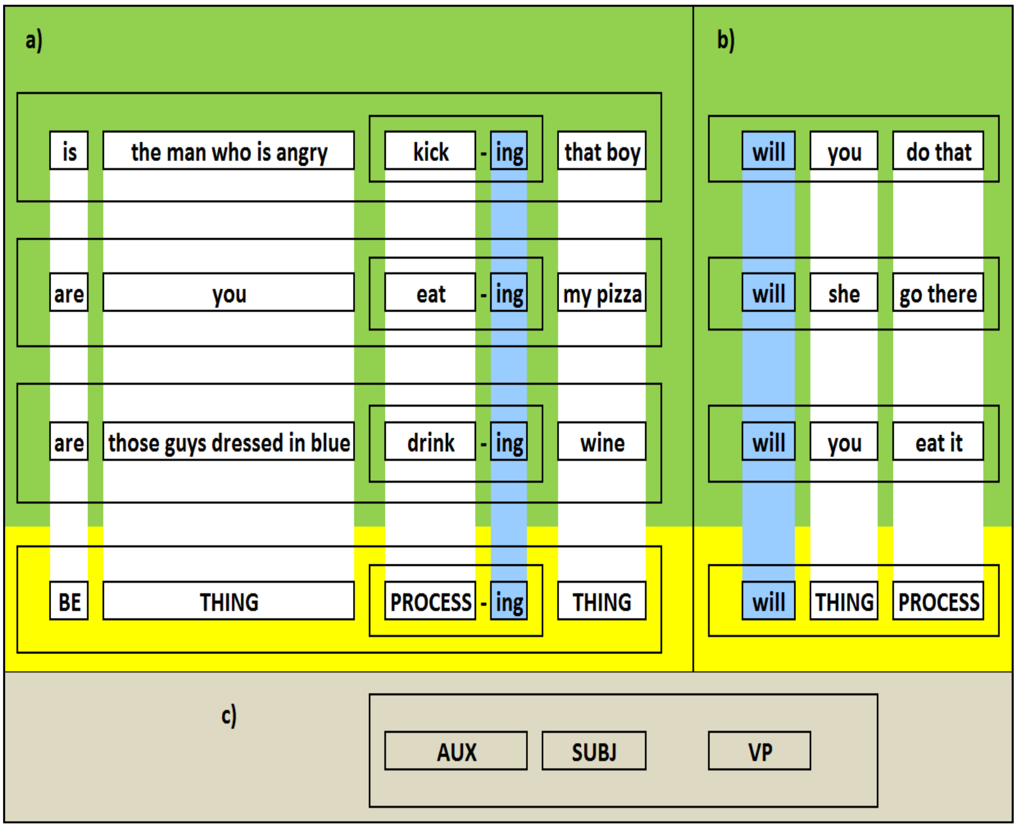


Figure 0.2: A UBA to learning syntactic questions in English. Concrete expressions from which schemas are inferred are in the green strip. Schemas are in the yellow strip. The grey strip (c) indicates that semantic generalisations (yellow strip) may gradually develop into more adultlike (possibly syntactic) ones. Slot formation (generalisations) is highlighted in white. Recurring lexical material is highlighted in blue.

0.3. THE CURRENT RESEARCH

UB researchers have overwhelmingly focussed on the English language (Tomasello, 2000a, 2003) and, consequently, their theoretical hypotheses are mostly based on results from studies that investigated the linguistic development of English-speaking children. English is a fairly exotic language, as it presents many peculiar characteristics, such as a poor morphological system and a high reliance on word order that

> "failed to appear to have equivalent strength in any other language"
>
> (Bates & MacWhinney, 1987, p. 172).

It is therefore not clear how well UB results would generalise to other languages (but see chapter 4). This research enquires into whether a UBA can account for the acquisition of Italian, a language with a rich morphological system and whose flexible word order is discourse-driven.

In order to enquire into the cross-linguistic validity of a UBA to LA, two research questions are posed:

a) Can Italian-speaking children's early language be accounted for in terms of lexically-specific units acquired from the concrete language that children themselves have previously experienced?

b) To what extent can Italian-speaking children be said to rely on (have mastered) fully-schematic constructions/patterns?

The first research question is investigated by means of a longitudinal study which enquires into whether the spontaneous production (*I eat pasta*) of an Italian-speaking two-year-old can be accounted for in terms of lexically-specific units (*EATER-eat-THING_EATEN*) instantiated in the specific strings (*we eat pizza*; *you eat soup*) he previously encountered. This question directly taps into learnability – (i) in 0.1 – as it investigates the extent to which the mother tongue can be learnt by drawing lexically-bound generalisations from the input.

The second question is investigated through an experimental study which sheds light on the development of children's linguistic productivity from the third to the fifth year of life and the extent to which it can be said to be fully-schematic (or abstract). Hence, it provides an insight into how children's morpho-syntactic competence develops over time (ontogeny, point iii in 0.1).

0.4. STRUCTURE OF THE THESIS

The current work is divided into two volumes: Volume I, which contains the main text, and Volume II, which contains the appendix.

In Volume I, after the current introduction, Part I begins by giving an overview of some of the relevant morpho-syntactic characteristics of Italian (chapter 1). The following chapters provide the reader with an overview of UBAs to language (chapter 2) and LA (chapter 3). Chapter 4 walks the reader through the rationale behind the design of the research. Part II and Part III present methodology, results and analyses of the longitudinal and experimental studies, respectively. Part IV ends the thesis with a unified discussion of the results of both studies (Chapter 16) and some general conclusions (chapter 17).

Part V of the thesis is contained in the second volume and comprises three Appendices. Appendix_I reports all examples, tables and figures of the main text. This is meant to help the reader when references to figures or examples that belong to previous chapters are made. By using Appendix_I, the reader can look at previous figures and examples without having to interrupt the reading process sifting through pages. Appendix_II and Appendix_III present relevant further information on the longitudinal and experimental studies respectively, which could not be included in the main text for reasons of space.

Page intentionally left blank

Part I

Background

Page intentionally left blank

1.

THE ITALIAN LANGUAGE

The following chapter walks the reader through some of the main phenomena that set Italian apart from English. This brief overview is by no means an exhaustive account of the morpho-syntactic phenomena described, let alone of the main differences between the two languages. Rather, this chapter provides the non Italian-speaker with some background, which is likely to facilitate a more thorough understanding of the design, results and analyses of the longitudinal and experimental studies to be discussed in part II and III, respectively.

1.1. MORPHOLOGICAL SYSTEM

Italian is a highly inflected language in which both nouns and verbs always present some kind of affix carrying specific grammatical information. Bare root forms do not occur.

1.1.1. ARTICLES, NOUNS AND ADJECTIVES

In Italian every noun is either masculine or feminine in gender; 75% of Italian nouns have a singular form ending in either *–a* when feminine (*sedi-a*, "chair"), or *–o* when masculine (*giorn-o*, "day") in gender. Their plurals are formed with the vowels *–e* (*sedi*-e "chairs") and *–i* (*giorn-i*; days), respectively. Most remaining nouns end with *–e* and can belong to either gender (*mes-e* "month" [masc]; *art-e*; "art" [fem]) and their plural forms end in *–i* (*mes-i* "months"; *art-i* "arts") (Tartaglione, 1997)[3]. Gender and number information is also carried by articles and adjectives. Some adjectives, such as *bello* "pretty/nice", carry both gender and number information, whereas others, such as *grande* "big", carry only number information (table 1.1.) (Tartaglione, 1997).

[3] The pattern is not without exceptions. For example, ***mano*** "hand" is feminine in gender, even though it ends in *–o*.

Table 1.1: Types of adjectives in Italian.

type of adjective	SG		PL	
	M	F	M	F
only SG vs PL distinction	grand-e big-SG		grand-i big-PL	
SG vs PL and MASC vs FEM distinction	bell-o nice-M.SG	bell-a nice-F.SG	bell-i nice-M.PL	bell-e nice-F.PL

Adjectives, articles and nouns agree in gender and number (fig. 1.1).

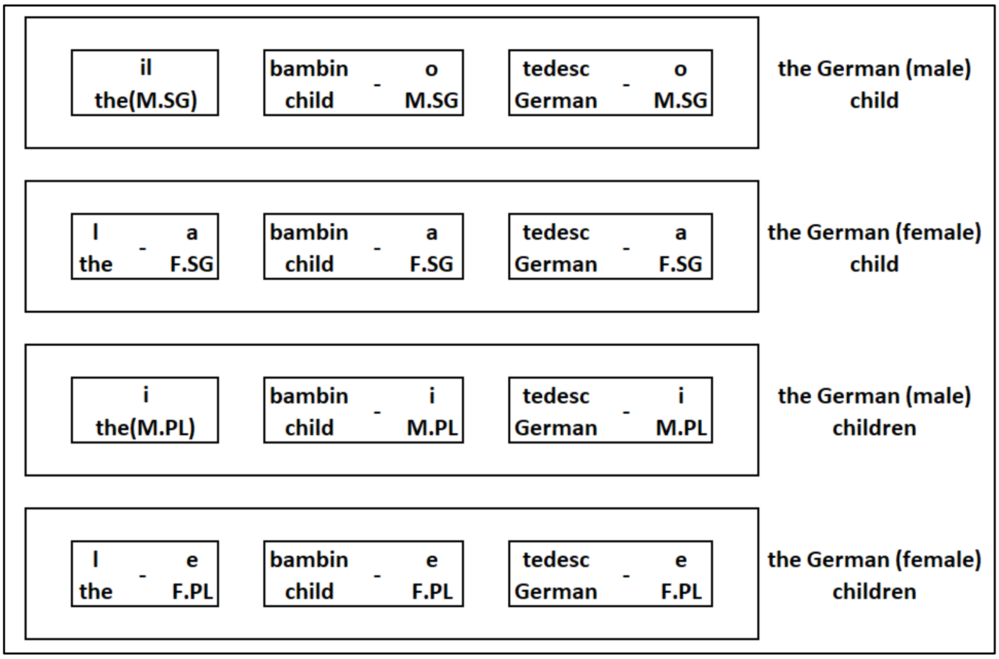


Figure 1.1: article-adjective-noun agreement in Italian and the regular gender-number markers (on adjectives and nouns) -*o*(M.SG), -*i*(M.PL), -*a*(F.SG), -*e*(F.PL).

1.1.1.1. Distributional Properties of Adjectives

Adjectives denoting a distinctive quality (nationality, material, shape) have post-nominal position (1). Adjectives denoting inherent qualities (*good*, *honest*, *evil*) can either precede or follow the noun (Russo, 1954). However, the position of an adjective depends more on the semantic function the adjective has for the speaker, rather than on its semantics tout court. In post-nominal position, adjectives indicate a sub-class of their heads (2b). In pre-nominal position, they denote a (perceived) inherent quality of them (2a) (Nespor, 2001). Hence, (1b) is possible if *sharp* is construed as an inherent quality of the blade.

1. a) *Una lama affilata*
   un-a *lam-a* *affilat-a*
   a-F.SG blade-F.SG sharp-F.SG

   b) *?? Un' affilata lama*
   ?? un affilat-a lam-a
   ?? a sharp-F.SG blade-F.SG

   "A sharp blade."

2. a) *I miei vecchi cappelli*
   i mi-ei vecch-i cappell-i
   the(M.PL) my-M.PL old-M.PL hat-M.PL
   "My old hats (my hats, which are old)."

   b) *I miei cappelli vecchi*
   i mi-ei cappell-i vecch-i
   the(M.PL) my-M.PL hat-M.PL old-M.PL
   "My old hats (the sub-class of my hats that are old)."

When an adjective is modified by an adverb or has complement(s), the order ADJ-NOUN is not possible (3) (Russo, 1954; Nespor, 2001).

3. a)* *Una molto bella ragazza*
   * *un-a molto bell-a ragazz-a*
   * a-F.SG very pretty-F.SG girl-F.SG

   b) *Una ragazza molto bella*
   un-a ragazz-a molto bell-a
   a-F.SG girl-F.SG very pretty-F.SG

   "A very pretty girl."

Italian can also modify adjectives through morphological synthetic means (Nespor, 2001; Scalisse, 2001). For instance, the affix *–issim-* is attached to the root of the adjective and is used to derive the superlative form (see Guasti, 2001). When an adjective is modified synthetically, it can either precede or follow its noun (4) (Nespor, 2001).

4. a) *Una ragazza bellissima*
   un-a ragazz-a bell-issim-a
   a-F.SG girl-F.SG pretty-very-F.SG

   b) *Una bellissima ragazza*
   un-a bell-issim-a ragazz-a
   a-F.SG pretty-very-F.SG girl-F.SG

   "A gorgeous girl."

Possessives (*mio* "my"; *suo* "her/his") agree in gender and number with the ownee and not with the owner. Hence, (5a) and not (5b) is the appropriate way of referring to Claire's brother.

5. a) *Suo fratello*
   su-o *fratell-o*
   his/her-M.SG brother-M.SG

   b) **Sua fratello*
   su-a *fratell-o*
   his/her-F.SG brother-M.SG

   "Her brother."

1.1.1.2. Diminutive and endearment affixes

Italian can express certain characteristics, such as size, both synthetically (inflectional morphology) and analytically (adj + noun sequences). Thus, in order to speak about a little table, Italian can use either analytical (6b-c) or synthetic (6a) forms.

6. a) *Il tavolino*
   il *tavol-in-o*
   the(M.SG) table-little(DIM)-M.SG

   b) *Il piccolo tavolo*
   il *piccol-o* *tavol-o*
   the(M.SG) little-M.SG table-M.SG

   c) *Il tavolo piccolo*
   il *tavol-o* *piccol-o*
   the(M.SG) table-M.SG little-M.SG

   "The little table."

Particularly relevant for this research are the diminutive affix *-in-* "little/small" (6a) and the endearment affix *–ett-* "little/cute" (*cas-ett-a* "home-little.cute(ENDR)-F.SG"), as they are very frequent in the longitudinal corpus collected (being typical of motherese).

1.1.2. VERBAL MORPHOLOGY

Italian verbs always agree with their subject(s) and carry information of aspect, mood, tense, person and number. Past participles may additionally carry gender information.

1.1.2.1. Inflectional classes.

According to traditional descriptive grammars (e.g. Barbieri, 1971), Italian verbs are grouped into three inflectional classes, called **conjugations**. Each conjugation is characterised by a different thematic vowel (TV, henceforth) and differs in terms of regularity, productivity and size (table 1.2) (Say & Clahsen, 2002; Barbieri, 1971)[4]. The number (and proportion) of verbs each class has depends on whether low frequency verbs are considered or not (table 1.3).

Table 1.2: the three conjugational classes of Italian.

Conjugation	Tematic Vowel TV	Example	Proportion of irregular verbs
1st Conjugation (or Conjugation I)	a	am - a - re love - TV - INF	four verbs and their derivatives
2nd Conjugation (or Conjugation II)	e	tem - e - re fear - TV - INF	95%
3rd Conjugation (or Conjugation III)	i	sent - i - re hear - TV - INF	10%

Table 1.3: the distribution of first, second and third conjugation verbs in Italian, according to different sources.

Primary sources used for this Research	Secondary sources on which primary sources draw	no. of verbs in each conjugation		
		Conjugation I	Conjugation II	Conjugation III
Barbieri (1971)	*Nuovo Vocabolario Zingarelli*	13,465	1,556	1,966
Orsolini & Marslen-Wilson (1997)	Dressler & Thornton (1991)	more than 3000	less than 400	more than 500
Say & Clahsen (2002)	De Mauro et al. (1993) and Bortolini et al. (1971)	1,709	403	238

Conjugation I presents the TV *–a-* and is the largest and most productive class; loans, onomatopoeias, denominals and neologisms are assigned to this class. Moreover, Italian has many verb-modifying affixes that assign a modified verb to Conjugation I, irrespective of its original class. For instance, *dormicchiare*

---

[4] Throughout this thesis, whenever the Thematic Vowel (TV) is included in the morphological glosses, it will be reported as *TV*. The conjugation class to which the TV belongs will be coded as *TV(conj.I)*, *TV(conj.II)* and TV(*conj.III)* when the morphological glosses are in the main text. When such information is reported in the figures, it is coded as *TV(1st.conj)*, *TV(2nd.conj)* and *TV(3rd.conj)*.

"to snooze" is a Conjugation I verb that derives from Conjugation III *dorm-i-re* "to sleep" (Say & Clahsen, 2002). All Conjugation I verbs but four (and their derivatives) are regular (Orsolini & Marslen-Wilson, 1997).

Conjugation II verbs have *–e-* as TV and infinite form ending in *–ere* (*ved-e-re* "to see"). Conjugation II is not productive and 95% of its verbs are irregular (Say & Clahsen, 2002; Orsolini & Marslen-Wilson, 1997).

Conjugation III verbs have the TV *–i-*. Conjugation III is still productive as de-nominal and de-adjectival inchoative verbs are assigned to this class (Say & Clahsen, 2002). Roughly 10% of Conjugation III verbs, and their derivatives, are irregular (Say & Clahsen, 2002; Orsolini & Marslen-Wilson, 1997).

1.1.2.2. Regular processes

Italian verbs are traditionally analysed as being formed by root + TV + aspect-mood-tense-person-number markers. The combination of root and TV forms the so called verb **stem**. Roots are bound morphemes (Orsolini & Marslen-Wilson, 1997) and specific verb forms are obtained by combining roots or stems with various kinds of affixes. Sometimes the bare root combines with aspect-mood-tense-person-number markers (table 1.4) and sometimes the latter combine with the stem (1.5). Aspect-mood-tense affixes can also be inserted between person-number markers and either the stem (table 1.6) or the bare root (table 1.7).

Table 1.4: Conjugating verbs in Italian: bare root + aspect-mood-tense-person-number affixes.

Verb	Tense, mood and aspect	Conj	root	aspect-mood-tense-person-number affix	resulting form	meaning
amare (to love)	ind prs	I	am	o	amo	I love
amare (to love)	ind prs	I	am	ano	amano	they love
temere (to fear)	ind prs	II	tem	o	temo	I fear
temere (to fear)	ind prs	II	tem	ono	temono	they fear
sentire (to hear)	ind prs	III	sent	o	sento	I hear
sentire (to hear)	ind prs	III	sent	ono	sentono	they hear
amare (to love)	subjunctive prs	I	am	i	ami	(that) she/he love
temere (to fear)	subjunctive prs	II	tem	a	tema	(that) she/he fear
sentire (to hear)	subjunctive prs	III	sent	a	senta	(that) she/he hear
amare (to love)	imp	I	am	a	ama!	(you) love!
temere (to fear)	imp	II	tem	i	temi!	(you) fear!
sentire (to hear)	imp	III	sent	i	senti!	(you) hear!

Table 1.5: Conjugating verbs in Italian: stem + aspect-mood-tense-person-number affixes.

Verb	Tense, mood and aspect	Conj	stem (root + tv)	aspect-mood-tense-person-number affix	resulting form	meaning
amare (to love)	past-definite (or preterite)	I	am + a	sti	amasti	you loved
amare (to love)	past-definite (or preterite)	I	am + a	stress	amò	she/he loved
temere (to fear)	past-definite (or preterite)	II	tem + e	sti	temesti	you feared
temere (to fear)	past-definite (or preterite)	II	tem + e	stress	temè	she feared
sentire (to hear)	past-definite (or preterite)	II	sent + i	sti	sentisti	you heard
sentire (to hear)	past-definite (or preterite)	III	sent + i	stress	sentì	she/he heard
amare (to love)	inf	I	am + a	re	amare	to love
temere (to fear)	inf	II	tem + e	re	temere	to fear
sentire (to hear)	inf	III	sent + i	re	sentire	to hear

Table 1.6: Conjugating verbs in Italian: stem + aspect-mood-tense affixes + person-number affixes.

Verb	Tense, mood and aspect	Conj	stem (root + tv)	apsect-mood-tense	person-number affix	resulting form	meaning
amare (to love)	ind imperf	I	am + a	v	o	amavo	I used to love
temere (to fear)	ind imperf	II	tem + e	v	o	temevo	I used to fear
sentire (to hear)	ind imperf	III	sent + i	v	o	sentivo	I used to hear

Table 1.7: Conjugating verbs in Italian: root + aspect-mood-tense affixess + person-number affixes.

Verb	Tense, mood and aspect	Conj	root	aspect-mood-tense specific infix	person-number affix	resulting form	meaning
amare (to love)	fut	I	am	er	emo	ameremo	we will love
temere (to fear)	fut	II	em	er	emo	temeremo	we will fear
sentire (to hear)	fut	III	sent	ir	emo	sentiremo	we will hear

1.1.2.3. Irregular verbs

Irregular verbs deviate from the patterns outlined above in different ways. For example, *andare* "to go" (Conjugation I) presents a phenomenon called **suppletivism**; namely the verb uses a different root with strong forms[5] (emboldened in table 1.8)

Table 1.8: irregular verbs, suppletivism: present indicative of *andare* "to go".

person-number	stem/root	aspect-mood-tense-person-number	resulting form	meaning
1.SG	**vad**	**o**	**vado**	**I go**
2.SG	**va**	**i**	**vai**	**you go**
3.SG	**va**	**Ø**	**va**	**s/he goes**
1.PL	and	iamo	andiamo	we go
2.PL	and	ate	andate	you go
3.PL	**va**	**nno**	**vanno**	**they go**

As previously mentioned, 95% of Conjugation II verbs are irregular. Irregularities mainly (but not only) occur with past participle and past-definite (i.e. preterite). Roughly 74% of irregular Conjugation II verbs belong to a subgroup of verbs (*root-change verbs*), which share certain morpho-phonological characteristics that make their irregularities partly predictable (see

[5] Broadly speaking, strong forms are those forms whose primary stress is on the stem (Barbieri, 1971, p. 197).

Orsolini & Marslen-Wilson, 1997). These verbs have roots ending in *–nd-* (*prendere*, "to take"), *vowel + -d-* (*ridere*, "to laugh"), *-t-* (*mettere*, "to put") or *-g-* (*leggere*, "to read") that become either *–ss-* or *–s-* with first and third singular persons and third plural person of the past-definite (tab. 1.9).

Table 1.9: indicative present and past-definite (preterite) of Conjugation II *root-change verbs*.

Verb	Tense, mood and aspect	root	aspect-mood-tense-person-number	resulting form	meaning
prendere (to take)	ind prs	prend	iamo	prend-iamo	we take
prendere (to take)	past-definite (peterite)	pres	i	pres-i	I took
ridere (to laugh)	ind prs	rid	iamo	rid-iamo	we laugh
ridere (to laugh)	past-definite (peterite)	ris	i	risi	I laughed
mettere (to put)	ind prs	mett	iamo	mett-iamo	we put
mettere (to put)	past-definite (peterite)	mis	i	mis-i	I put
leggere (to read)	ind prs	legg	iamo	legg-iamo	we read
leggere (to read)	past-definite (peterite)	less	i	less-i	I read

Root ending however, is not always a reliable cue of a verb belonging to this subclass. For instance, *no-root-change verbs*, such as *vendere* "to sell" (root ending *–nd-*), can either follow the default, regular pattern, or add one of the interfixes *–tt-* or *–st-* between stem and person-number affixes (table. 1.10) (Orsolini, Fanari & Bowles, 1998)[6].

[6] According to English-speaking linguists, an *interfix* is a morpheme that does not carry any meaning and that is inserted *within* the root. As discussed above (see also table 1.10), some regular, no-root-change verbs of Conjugation II can have either a default and regular form (stem-marker; see (a) below) or a non-default form (see (b) below).

a) Vende(stem)-rono(PAST.2.PL)= venderono = they sold
b) Vende(stem)-tt(interfix)-ero(PAST.2.PL) = vendettero = they sold

However, in (b) the "interfix" *–tt-* has no meaning and is not used lest there be some kind of phonetic/phonological violation. It is very much a stylistic choice. Following some descriptions of Italian morphology in English articles (e.g. Orsolini et al., 1998), I refer to those morphemes as *interfixes*, even though they are technically inserted after the stem or root and not within the root.

Table 1.10: No-root-change verbs; past-definite (preterite).

No-root-change Verbs (*vendere* "to sell")						
Verb		stem (root + tv)	tense-specific interfix	person-number marker	resulting form	meaning
default option	vendere (to sell)	vend + e	-	stress	vendé	s/he sold
non-default option	vendere (to sell)	vend + e	tt	e	vendette	s/he sold
default option	vendere (to sell)	vend + e	-	rono	venderono	they sold
non-default option	vendere (to sell)	vend + e	tt	ero	vendettero	they sold

1.1.2.4. Passato Prossimo

The experimental study elicits productivity with *passato prossimo* of regular Conjugation I verbs. The *Passato prossimo* is the most common past tense in varieties of Northern Italian and is formed in a similar manner to the English *present perfect*: by combining an auxiliary (*be* or *have*) with the past participle of the lexical verb. The regular past participle is formed by adding the affix *–t-* and gender-number regular markers[7] to the verb stem (table 1.11). Some Conjugation II *no-root-change verbs* (e.g. *esistere* "to exist") have Conjugation III past participles (*esist-i-to*, not **esist-u-to*).

Table 1.11: Forming regular participles in Italian.

Verb	Conj	root	TV	particple marker	number-gender suffix	resulting form	meaning
amare (to love)	I	am	a	t	o/i	ama-t-o/i	loved-M.SG/M.PL
amare (to love)	I	am	a	t	a/e	am-a-t-a/e	loved-F.SG/F.PL
temere (to fear)	II	tem	u*	t	o/i	tem-u-to/i	feared-M.SG/M.PL
temere (to fear)	II	tem	u*	t	a/e	tem-u-t-a/e	feared-F.SG/F.PL
sentire (to hear)	III	sent	i	t	o/i	sent-i-t-o/i	heard-M.SG/M.PL
sentire (to hear)	III	sent	i	t	a/e	senti-t-a/e	heard-F.SG/F.PL

* *-u-* is the allomorph of conjugation II's tv *–e-*.

Conjugation II *root-change verbs* form the past participle by either using the same changed root used for the past-definite (e.g. *ridere* "to laugh"; see table 1.12) or by attaching *–st-* or *–t-* to another modified root (e.g. *piangere* "to cry"; see table 1.12). Other irregular verbs follow even more idiosyncratic patterns.

[7] *-o* "M.SG", *-i* "M.PL", *-a* "F.SG" and *–e* "F.PL".

Table 1.12: Past participle of irregular Conjugation II verb.

process	verb	regular root	modified root	-/st/ or -/t/	gender-number suffix (M.SG)	resulting form	meaning
same root of past-definite (preterite)	ridere	rid	ris	-	o	riso	laughed
other root + -/st/ or -/t/	piangere	piang	pian	t	o	pianto	cried

When *essere* "to be" is the auxiliary verb, the past participle and subject(s) agree in gender and number (7).

7. *Le ragazze sono cadute*
   *l-**e*** *ragazz-**e*** *sono* *cad-u-t-**e***
   the-**F.PL** girl-**F.PL** be(PRS.3.PL) fall-TV(conj.II)-PTCP-**F.PL**
   "The girls fell."

When *avere* "to have" is the auxiliary, there is no subject-participle agreement and the default M.SG morpheme *–o* is used (8), unless the direct object (DO, henceforth) is a 3.ACC clitic pronoun. In this case, there **must** be gender-number agreement between the participle and the clitic pronoun (9).

8. *Sara ha spostato la sedia*
   Sara *ha* *spost-a-t-**o***
   Sara has move-TV(conj.I)-PTCP-**M.SG(default)**
   *l-**a*** *sedi-**a***
   the-**F.SG** chair-**F.SG**
   "Sara moved the chair."

9. a) *Le ho mangiate*
   *l-**e*** *ho* *mangi-a-t-**e***
   clitic.3.ACC.-**F.PL** have(PRS.1.SG) eat-TV(conj.I)-PTCP-**F.PL**

   b) **Le ho mangiato*
   l-e*** *ho* *mangi-a-t-**o***
   *clitic.3.ACC-**F.PL** have(PRS.1.SG) eat-TV(conj.I)-PTCP-**M.SG**

   "I ate them."

1.2. NULL SUBJECT

Italian allows sentences not to overtly express their subjects and Italian speakers' sentences are overwhelmingly (70%) subjectless (Lorusso, Caprini & Guasti, 2005). Null subject sentences are possible when the subject can be

> "unambiguously recovered"
>
> (Serratrice, 2005, p. 442).

This includes, but is not limited to, meteorological verbs (10), cases in which the subject is inferable by either the linguistic or extra-linguistic context (11) and first and second persons (12) that are mostly deictic and

> "essentially active by default"
>
> (Serratrice, 2005, p. 443)

(see also Benincà with Salvi & Frison, 2001).

10. *Nevica molto*
    nevic-a *molto*
    snow-PRS.3.SG a.lot
    "It snows a lot."

11. *Guarda! Non è stanca qui?*
    guard-a! *non* *é* *stanca-a* *qui?*
    look-IMP.2.SG! not is tired-F.SG here?
    "Look! Isn't she tired, here?"
    Context: while Looking at a girl's picture.
    (Serratrice, 2005, p. 442).

12. *Vado via*
    vad-o *via*
    go-PRS.1.SG away.
    "I go away."

Subjects are overtly uttered when speakers need to disambiguate them among other possible subjects (13b), activate them as new elements in the discourse (14) and emphasise them contrastively (15) (Serratrice, 2005).

13. a) *Cosa hanno fatto Chiara e Marco ieri?*

cosa	*hanno*	*fatto*	*Chiara*	*e*
what	have(PRS.3.PL)	done	Chiara	and
Marco	*ieri?*			
Marco	yesterday?			

"What did Chiara and Marco do yesterday?"

b) *Sono andati alla festa, ma lei è andata a casa presto*

sono	*and-a-t-i*	*a=(l)l-a*		
be(PRS.3.PL)	go-TV(conj.I)-PTCP-M.PL	to=the-F.SG		
fest-a,	*ma*	*lei*	*é*	*anda-t-a*
party-F.SG,	but	she	is	go-TV(conj.I)-PTCP-F.SG
a	*cas-a*	*presto*		
to	home-F.SG	early		

"They went to the party, but she went home early."

14. *Dopo la festa, Paolo mi ha dato un passaggio a casa*

dopo	*l-a*	*fest-a,*	*Paolo*	*mi*
after	the-F.SG	party-F.SG,	Paolo	clitic.1.SG.DAT
ha	*d-a-to*	*un*		
has	give-TV(conj.I)-PTCP	a(M.SG)		
passaggi-o	*a*	*cas-a*		
lift-M.SG	to	home-F.SG		

"After the party, Paolo gave me a lift home."
(From Serratrice, 2005, p. 444)

15. *IO, ho pulito dopo la festa (non tu)!*

IO,	*ho*	*pul-i-to*	*dopo*
I,	have(PRS.1.SG)	clean-TV(conj.III)-PTCP	after
l-a	*fest-a*	*(non*	*tu)!*
the-F.SG	party-F.SG	(not	you)!

"***I***, cleaned after the party (not you)!"

1.3. CLITIC PRONOUNS

1.3.1. CLITIC AND TONIC PRONOUNS

The Italian pronominal system distinguishes between two sets of pronouns; **tonic** (also called free or **strong** pronouns) and **clitic** (table 1.13).

Table 1.13: the Italian pronominal system (no possessives). Adapted from Cordin and Calabrese (2001).

Type of Pronoun	case	Singular Forms				Plural Forms			
		1	2	3 masc	3 fem	1	2	3 masc	3 fem
TONIC (or Strong) PRONOUNS	nom	io	tu	lui/egli**	lei/ella**	noi	voi	loro/ essi**	loro/esse**
	acc	me	te	lui/esso**	lei/essa**	noi	voi	loro	loro
CLITIC PRONOUNS	acc	mi	ti	lo	la	ci	vi	li/gli	le
	dat	mi	ti	gli/ci	le/gli/ci	ci	vi	gli/ loro	gli/ loro

** formal language

When a strong (or tonic) pronoun is used for an indirect case, its accusative form is combined with the appropriate preposition: *te* "you(2.SG.ACC)", *a te* "to you".

Strong pronouns can occur in any sentence position and have independent accent. The real-life element to which they refer is new or not clearly activated. They are often used to express a contrastive, new or emphasised element, as *lei* "she" in (13b) (Cordin & Calabrese, 2001).

Clitic pronouns have neither their own accent nor an independent position. They always attach to a verb and their position depends on the form of the verb. The real-life element to which they refer is given, already activated (Cordin & Calabrese, 2001). The question in (16a) activates *coltelli* "knives" and makes them *discourse topic.* A clitic pronoun can therefore refer to *i coltelli* "the knives" because these can be assigned *old/given* value (16b).

16. a) *Dove sono **i coltelli**?*

dove	*sono*	***i***	***coltell-i***
where	be(PRS.3.PL)	**the(M.PL)**	**knife-M.PL**

"Where are the knives?"

b) ***Li** ho presi io*

l-i	*ho*	*pres-i*	*io*
clitic.3.ACC-M.PL	have(PRS.1.SG)	taken-M.PL	I

"I took them."

1.3.2. CLITICS' DISTRIBUTIONAL PROPERTIES

Clitics have fixed positions depending on whether they are combined with a [+FINITE] or [-FINITE] verb. In the former case they take pre-verbal position (17); in the latter (e.g. infinitives, **positive** imperatives) they merge with the verb in post-verbal position (18-19) (Cordin & Calabrese, 2001). In the case of **negative** imperatives, however, clitics can have either pre or post-verbal position (20a-b).

17. *Lo vedo*

l-o	*ved-o*
clitic.3.ACC-M.SG	see-PRS.1.SG

"I see it/him."

18. *Mangiarlo*[8]

*mangi-a-r=**l-o***
eat-TV(conj.I)-INF=**clitic.3.ACC-M.SG**
"To eat it/him."

19. *Mangialo!*

*mangi-a=**l-o**!*
eat-IMP.2.SG=**clitic.3.ACC-M.SG**
"Eat it/him!"

20. a) *Non lo mangiare!*

non	***l-o***	*mangi-a-re!*
not	**clitic.3.ACC-M.SG**	eat-TV(conj.I)-INF/IMP.2.SG

b) *Non mangiarlo!*

non	*mangi-a-r=**l-o***
not	eat-TV(conj.I)-INF/IMP.2.SG=**clitic.3.ACC-M.SG**

"Don't eat it/him!"

[8] According to Lehmann's (1982) guidelines, the symbol "=" indicates morpho-phonological merging (i.e. cliticisation). In (18) it indicates that the clitic merges with the verb.

If clitics appear in main clauses with a [+FINITE] auxiliary, they attach to AUX[9] in preverbal position (21) (Cordin & Calabrese, 2001, p. 565):

21. *Mi ha dato il tuo libro*
    mi *ha* *d-a-to* *un*
    clitic.1.SG.DAT has give-TV(conj.I)-PTCP a(M.SG)
    libr-o
    book-M.SG.
    "(S/he) gave me a book."

When clitics belong to a subordinate whose verb has [-FINITE] form, there are two options. They can either merge with the [-FINITE] verb of the subordinate clause (22a; post-verbal position) or attach to the [+FINITE] verb of the main clause (22b; pre-verbal position; Cordin & Calabrese, 2001).

22. a) *Marco vuole mangiarlo!*
    Marco *vuol-e* *mangi-a-r=**l-o***
    Marco want-PRS.3.SG eat-TV(conj.I)-INF=**clitic.3.ACC-M.SG**

    b) *Marco lo vuole mangiare*
    Marco ***l-o*** *vuol-e*
    Marco **clitic.3.ACC-M.SG** want-PRS.3.SG
    mangi-a-re
    eat-TV(conj.I)-INF.

    "Marco wants to eat it/him."

    (Cordin & Calabrese, 2001, p. 587)

If the subordinate has a [+FINITE] verb, the clitic cannot attach to the verb of the main clause (23a-b).

[9] AUX, such as *essere* "to be" and *avere* "to have". Modals are discussed later.

23. a) *Carlo vuole che tu lo prenda*
Carlo vuol-e che tu
Carlo want-PRS.3.SG that you(2.SG.NOM)
l-o *prend-a.*
clitic.3.ACC-M.SG take-SBJV.PRS.2.SG

b) * *Carlo lo vuole che tu prenda*
Carlo ***l-o*** *vuol-e che*
Carlo **clitic.3.ACC-M.SG** want-PRS.3.SG that
tu prend-a
you(2.SG.NOM) take-SBJV.PRS.2.SG

"Carlo wants you to take it/him."

(Cordin & Calabrese, 2001).

No element can occur between verb and clitic (24a-b), unless it is another clitic (25)[10] (see Cordin & Calabrese, 2001).

24. a) *Lo voglio disperatamente*
l-o *vogli-o disperata-mente*
clitic.3.ACC-M.SG want-PRS.1.SG desperate-ly

b)* *Lo disperatamente voglio*
****l-o*** *disperata-mente vogli-o*
clitic.3.ACC-M.SG desperate-ly want-PRS.1.SG

"I desperately want it/him."

25. *Glielo dico*
gli=(e)=l-o *dic-o*
clitic.3.DAT.M.SG=(e)=clitic.3.ACC-M.SG say-PRS.1.SG
("I say him/it to him.")
"I say that to him."

1.3.3. CLITICS AND MORPHOLOGICAL AGREEMENT

When *passato prossimo* constructions are used, accusative clitics might agree in gender and number with past participles. Clitic-past_participle agreement is compulsory only with third persons (9), whereas it is optional with first and second persons (26) (Cordin & Calabrese, 2001).

[10]This does not apply to *loro* "to them"; it always has post-verbal position and other elements can occur between it and its verb (Cordin & Calabrese, 2001).

26. a) *Sara ci ha vist**o***

Sara	*ci*	*ha*	*vis-t-**o***
Sara	clitic.1.PL.ACC	has	see-PTCP-**M.SG**

b) *Sara ci ha vist**e***

Sara	*ci*	*ha*	*vis-t-**e***
Sara	clitic.1.PL.ACC	has	see-PTCP-**F.PL**

"Sara saw us."
Context: *ci* "us" refers to two or more girls.

1.3.4. RESUMPTIVE CLITICS AND THEMATIC ROLES

Clitics often play a fundamental role in the interpretation of sentences as resumptive clitics are frequently essential cues to thematic roles. Italian is fairly free with respect to word order (WO, henceforth) and elements are often "dislocated" (see 1.4). Agreement between dislocated element and co-indexed resumptive clitic makes thematic roles easily inferable, particularly when it comes to *AGENT vs. PATIENT* distinctions.

In (27) the two NPs (*il topo* "the mouse"; *la volpe* "the fox") are both in preverbal position and singular in number. Thus, neither WO nor verb morphology (PRS.3.SG) can disambiguate thematic roles. The puzzle is solved by clitic-NP agreement. *La volpe* "the fox" (F.SG) cannot be co-indexed with the clitic *lo* "clitic.3.ACC.M.SG", as they have different gender. Conversely, *il topo* "the mouse" and *lo* "clitic.3.ACC.M.SG" agree in gender and number (M.SG). Hence, *il topo* and *lo* are co-indexed; *lo's* accusative case makes it clear that *il topo* is the direct object (DO). By exclusion, *la volpe* is the subject.

27. *Il topo$_i$, la volpe lo$_i$ mangia*

[il	*top-**o]**$_i$,*	*l-a*	*volp-e*
[the(**M.SG)**	mouse**-M.SG]**$_i$,	the-F.SG	fox(F.)-SG
*[l-**o**]$_i$*	*mangi-a*		
[clitic.3.ACC-**M.SG**]$_i$	eat-PRS.3.SG		

"The fox eats the mouse."

1.3.5. THE CLITIC PRONOUN "*SI*"

The reflexive clitic *si* can be translated into English passives in certain contexts. For instance, (28) (analysed in fig. 8.11-8.12; section 8.3.2) literally means that the roof spoiled itself. However, the roof has here to be intended as a passive subject that undergoes, rather than performs, the action. The same construction with a [+ANIMATE] subject (29) can have either a passive (*the gentleman got all dirty*) or a purely reflexive (*the gentlemen dirtied himself*) meaning, depending on how speaker and listener construe the scene.

28. *Il tetto si è rovinato tutto*
il *tett-o* *si* *è*
the(M.SG) roof-M.SG clitic.3.REFL is
rovin-a-t-o *tutt-o*
ruin-TV(conj.I)-PTCP-M.SG all-M.SG
"The roof got all spoiled."

29. *Il signore si è sporcato tutto*
il *signor-e* *si* *è*
the(M.SG) gentleman(M.)-SG clitic.3.REFL is
sporc-a-t-o *tutt-o*
dirty-TV(conj.I)-PTCP-M.SG all-M.SG
"The gentleman completely dirtied himself" or "The gentleman got all dirty".

The clitic *si* can also be used as an *indefinite subject,* which (mostly) maps onto a [+HUMAN] [+PLURAL] [-DEFINITE] entity (see Salvi, 2001a-b). In those cases, if *si* is the subject of a transitive verb and constructions with the past participle (e.g. passives) are used, the 3.SG verb is combined with a plural past participle (30). (30) means that an unspecified [+HUMAN] plural entity is often ignored by politicians.

30. *Si è spesso ignorati dai politici*
si *è* *spesso* *ignor-a-t-i*
clitic.IMPRS.NOM is often ignore-TV(conj.I)-PTCP-M.PL
da=i *politic-i*
by=the(M.PL) politician-M.PL
"People are often ignored by politicians."

When the (active) verb of an *indefinite subject "si"* is transitive, it agrees in gender and number with the DO and not with *si*. Seen from another angle (Salvi, 2001a), the DO becomes the subject and *si* merely indicates passivisation of the verb (31).

31. *Si mangiano i pomodori*
si *mangi-ano* *i* *pomodor-i*
clitic.IMPRS.NOM eat-PRS.3.PL the(M.PL) tomato-M.PL
"Tomatoes get/are eaten."

Importantly, the *impersonal subject "si"* often takes on a *deontic* meaning: the subject represents a model to follow (32) (Salvi, 2001a).

32. *Non si fa così*
non *si* *fa* *così*
not clitic.IMPRS.NOM does like.that
"That is not the way to do it" or "This is not to be done" or "That's not the way to behave".

1.4. WORD ORDER

Italian presents a very a flexible word order (WO, henceforth), which is the by-product of a discourse pragmatics-driven ordering of phrasal constituents.

1.4.1. A TOPIC-COMMENT (GIVEN-NEW) LANGUAGE

Italian mostly adopts a *topic-comment/given-new* order (Salvi, 2001a). What is considered the default WO (SVO) is, in reality, the felicitous combination of a subject which is both *topic* and *given*; and a VP which is both *comment* and *new*. As Salvi (2001a, p. 52) points out, swapping constituents around is often the device Italian speakers adopt to restore a *given-new*/*topic-comment* order in the discourse.

A sentence can be pragmatically marked and yet syntactically default (33) and vice versa (34) (Salvi, 2001a).

33. *I RAGAZZI rincorrono il cane*
i *RAGAZZ-I* *rincorr-ono* *il* *can-e*
the(M.PL) boy-M.PL chase-PRS.3.PL the(M.SG) dog(M)-SG
"They boys chase the dog."

34. *Lo ha cucinato la mamma*
l-o *ha* *cucin-a-t-o*
clitic.3.ACC-M.SG has cook-TV(conj.I)-PTCP-M.SG
l-a *mamm-a*
the-F.SG mum-F.SG
"Mum cooked him/it."

(34) may be found in a context where X asks information as to who cooked the turkey s/he is eating and Y replies with (34). The order is syntactically marked (oVS[11]), but it is pragmatically default. The *topic/given* is that someone has cooked the turkey. The *new/comment* is that the mother did.

[11] Henceforth, whenever reference to a Direct Object (DO) is made, a small *o* indicates an ACC clitic pronoun (as in *oV*), whereas capital *O* (as in VO) indicates any other kind of DO (NPs, strong pronouns, etc.).

(33) shows that Italian can adopt pragmatically marked orders (*NEW-given*) by phonetically stressing (indicated by CAPITALS) the element that is new. *I ragazzi* "the boys" is a new element in initial position. Yet, the sentence is syntactically unmarked (*S*VO).

A case in point of the strength of such reliance on *topic-comment/given-new* structure is the aversion of Italian native-speakers to indefinite subjects, even though those are perfectly grammatical (Bates & MacWhinney, 1987). Bates and MacWhinney (1987, p. 162) report that when asked to describe a picture of a monkey eating a banana, Italian two-year-olds are six/seven times more likely to describe the scene with *there is a monkey eating a banana* than their English-speaking peers, who go for *a monkey is eating a banana*. Subjects are strongly associated with *topic* function, whereas indefinite articles are more likely to convey new information. Indefinite subjects somehow violate *topichood* and create a conflict. Resorting to "*there* constructions" is a way of maintaining a *topic-comment/given-new* structure.

1.4.2. SWAPPING COSTITUENTS AROUND

Traditionally (see Benincà with al., 2001), sequences other than *(S)-V-O-X* are considered syntactically marked orders resulting from left and right dislocations of phrasal constituents. Orders other than *given-new* are pragmatically marked. The latter can be orders such as *NEW-given* and *given-NEW-given* (Salvi, 2001a). Capitals indicate that, for a *new* element to occur on the left of a *given* one, *PHONETHIC STRESS* on the former is needed.

Four phenomena are associated with the **dislocation of elements to the left**. Left Dislocation (LD), Hanging Topic (HT) and Anaphoric Ante-position (AA) assign *given/topic* value to the moved element(s); Focalisation (FOC) assigns contrastive value of NEW to the element it dislocates. Since AA is typical of oratorical speech and has an Old-Italian flavour attached to it (Benincà with al., 2001), it is not discussed here.

When an indirect object (IO), such as a receiver, is considered dislocated, the whole PP appears in the left periphery with both LD and FOC. In the former case (35a), the PP is assigned *given/topic* value. In the latter case (36a), it has contrastive value of NEW, is phonetically stressed and is normally followed by a brief pause that isolates it from the rest of the sentence (Benincà with al. 2001). LD can optionally recall the PP, but only through a clitic pronoun, which agrees in gender, number and case with the dislocated element (35b). No element can co-index with the focalised constituent (36b) (see Benincà with al., 2001).

35. LD
a) *A Marco, Sara dà un oracchiotto.*
a Marco, Sara dà un orsacchiott-o
to Marco, Sara gives a(M.SG) teddy.bear-M.SG.
"Sara gives Marco a teddy bear."

b) *A Marco$_i$, Sara gli$_i$ dà un oracchiotto.*
[a Marco]$_i$ Sara [gli]$_i$ dà un
[to Marco]$_i$, Sara [clitic.3.DAT.M.SG]$_i$ gives a(M.SG)
orsacchiott-o
teddy.bear-M.SG
"Sara gives Marco a teddy bear."

c) **Marco, Sara dà un oracchiotto.*
**Marco, Sara dà un orsacchiott-o .*
Marco, Sara gives a(M.SG) teddy.bear-M.SG
"*Sara gives a teddy bear Marco."

36. FOC
a) *A MARCO, Sara dà un oracchiotto.*
A MARCO, Sara dà un orsacchiott-o
to Marco, Sara gives a(M.SG) teddy.bear-M.SG.
"Sara gives a teddy bear TO MARCO."

b)* *A MARCO$_i$, Sara gli$_i$ dà un oracchiotto.*
**[A MARCO]$_i$ Sara [gli]$_i$*
[TO MARCO]$_i$, Sara [clitic.3.DAT.M.SG]$_i$
dà un orsacchiott-o
gives a(M.SG) teddy.bear-M.SG.
"Sara gives Marco a teddy bear."

HT can be thought of as a topic followed by a sentence about it (i.e. a *topic-comment* construction). The speaker mentions a topic[12] and then says something about it. When an IO is a HT there is no preposition stranding and the topicalised NP must be recalled (37a-b), because otherwise its relationship with the rest of the sentence (its syntactic role) would be unclear (37c). Recalling may (37a), but need not (37b), happen through clitic pronouns (which agree in gender and

[12] Either a *topic* that is old information (already active in the discourse) or a *new topic* the speaker activates.

number with the HT). In (37b) *Marco* is co-indexed with the PP *a quel bambino* “to that child”, which recalls it.

37. HT

a) Marco$_i$, Sara gli$_i$ dà un oracchiotto.
[Marco]$_i$, Sara [gli]$_i$ dà
[Marco]$_i$, , Sara [clitic.3.DAT.M.SG]$_i$ gives
un orsacchiott-o
a(M.SG) teddy.bear-M.SG.
“Sara gives Marco a teddy bear.”

b) *Marco$_i$, Sara dà un orsacchiotto a quel bambino$_i$*
[Marco]$_i$, Sara dà un orsacchiott-o
[Marco]$_i$, Sara gives a(M.SG) teddy.bear-M.SG
[a quel bambin-o]$_i$
[to that(M.SG) child-M.SG]$_i$
“Marco, Sara gives a teddy-bear to that child.”

c) **Marco, Sara dà un oracchiotto.*
Marco, Sara dà un orsacchiott-o .
Marco, Sara gives a(M.SG) teddy.bear-M.SG
“Sara gives Marco a teddy bear.”

When a DO is left-dislocated it can have either *NEW* (FOC) or *given/topic* (LD, HT) value. In the former case the dislocated element is phonetically stressed and cannot be recalled (39). In the latter case, recalling is compulsory (38). LD can recall the dislocated element only by means of accusative clitics. HT can do so through a variety of means (clitic and tonic pronouns, co-indexed NPs, demonstrative pronouns, etc). Thus, (38a) could be due to either LD or HT, whereas (38b) is unambiguously a case of HT, as *Marco* is recalled by the NP *quell’uomo cattivo* “that mean man”. Phonetic stress and recall are mutually exclusive: a non-stressed, dislocated DO without any co-indexed element recalling it is ungrammatical[13] (38c) and so is a stressed dislocated DO that is recalled (39b).

[13] Although this may be possible with AA (see Benincà with al., 2001).

38. a) LD and HT
Marcoi, tutti loi incolpano
[Marco]i, *tutt-i* *[l-o]i* *incolp-ano*
[Marco]*i*, all-M.PL [clitic.3.ACC-M.SG]*i* blame-PRS.3.PL
"Everybody blames Marco."

b) HT
[Marco]i, tutti incolpano quell'uomo cattivoi
[Marco]i, *tutt-i* *incolp-ano*
[Marco]*i*, all-M.PL blame-PRS.3.PL
[quell' *uom-o* *cattiv-o]i*
[that(M.SG) man-M.SG mean-M.SG]*i*
"Marco, everybody blames that mean man."

*c) *Marco, tutti incolpano*
Marco, *tutt-i* *incolp-ano*
Marco, all-M.PL blame-PRS.3.PL
"Everybody blames Marco."

39. a) *IL GATTO, ho buttato fuori*
IL *GATT-O,* *ho*
the(M.SG) cat-M.SG, have(PRS.1.SG)
butt-a-to *fuori*
throw-TV(conj.I)-PTCP out

b)* *[IL GATTO]i, loi ho buttato fuori.*
** [IL* *GATT-O]i,* *[l-o]i*
* [the(M.SG) cat-M.SG]*i*, [clitic.3.ACC-M.SG]*i*
ho *butt-a-t-o* *fuori*
have(PRS.1.SG) throw-TV(conj.I)-PTCP-M.SG out

"I threw THE CAT out."

Finally, Right Dislocation (RD), moves the **topic/given of the discourse to the right edge**[14]. RD mainly occurs with "out-of-the-blue" sentences, is highly context-dependent and mirrors speakers' assumptions. Resumptive clitics are optional and not compulsory (40) (Benincà with al., 2001).

[14] I ignore *Emarginazione* (see Cardinaletti, 2001; Benincà with al., 2001) for three reasons. Firstly, it is often difficult to tell it apart from RD (Benincà with al., 2001). Secondly, it is thought to be brought about by other dislocation processes (Benincà with al., 2001; Antinucci & Cinque, 1977). Finally, if one bears in mind that in Italian word order is flexible, such sentences are unproblematic to interpret (see Benincà with al., 2001).

40. *(Lo)ᵢ porto domani, il dolceᵢ*

([l-o] ᵢ) *port-o* *domani,*

([clitic.3.ACC-M.SG]ᵢ) bring-PRS.1.SG tomorrow,

[il *dolc-e]ᵢ*

[the(M.SG) dessert(M.)-SG]ᵢ

"Tomorrow I'll bring the dessert."

(Benincà with al., 2001, pp. 160-161)

The examples above show that Italian allows virtually any linear order of the major clausal constituents (VP, PP, etc.); the exact way in which constituents are ordered is discourse-driven. Importantly, the flexible ordering of major constituents described above with regard to simple clauses also applies to constituents in complex sentences, allowing them to cross clause-boundaries. The examples in (41) below illustrate some of the possible ways in which such flexibility can be exploited in complex sentences (from Antinucci & Cinque, 1977, note 2, pp. 143-144; colours indicate constituents).

41. a. *Quando ha detto, Giorgio, che avrebbe smesso di piovere, a voi?*

quando ha detto, *Giorgio,* *che* *avr-ebbe*

when has said, Giorgio, that have-COND.PRS.3.SG

smesso di piov-e-re, *a voi*

stopped to rain-TV(conj.II)-INF, to you(2.PL)

b. *Quando ha detto, Giorgio, a voi, che avrebbe smesso di piovere?*

quando ha detto, *Giorgio,* *a voi,* *che*

when has said, Giorgio, to you(2.PL) that

avr-ebbe *smesso di piov-e-re*

have-COND.PRS.3.SG stopped to rain-TV(conj.II)-INF

c. *Quando ha detto, a voi, che avrebbe smesso di piovere, Giorgio?*

quando ha detto, *a voi,* *che*

when has said, to you(2.PL) that

avr-ebbe *smesso di piov-e-re,*

have-COND.PRS.3.SG stopped to rain-TV(conj.II)-INF ,

Giorgio

Giorgio

"When did Giorgio tell you that it would stop raining?"

1.4.3. DEFAULT DISLOCATIONS: Psychological Verbs

Some psychological verbs, such as *piacere* "to like", have a [-ANIMATE] subject experienced by a [+ANIMATE] dative IO. The default syntactic order would be *S-V-IO* (42).

42. *I pomodori piacciono a Marco*
i pomodor-i piacci-ono a Marco
the(M.PL) tomato-M.PL like-PRS.3.PL to Marco
"Marco likes tomatoes."

Such an order creates a conflict between grammatical subject and semantic experiencer, as agents/experiencers are normally *topic* in the discourse. This conflict determines the fact that sentences like (43) – order *IO-V*-S - rather than (42) – order *S-V-IO* -, represent the most used pattern in current Italian (Benincà with al., 2001).

43. *A Marco piacciono i pomodori*
a Marco piacci-ono i pomodor-i
to Marco like-PRS.3.PL the(M.PL) tomato-M.PL
"Marco likes tomatoes."

In varieties of Italian spoken in Northern Italy, such a construction normally uses a resumptive dative clitic, which is co-indexed with the experiencer (44-45).

44. *[A Marco]i glii piacciono i pomodori*
[a Marco]i [gli]i piacci-ono
[to Marco]*i* [clitic.3.DAT.M.SG]*i* like-PRS.3.PL
i pomodor-i
the(M.PL) tomato-M.PL
"Marco likes tomatoes."

45. *A me, mi piacciono i pomodori.*
[a me]i [mi]i piacci-ono i
[to me]*i* [clitic.1.SG.DAT]*i* like-PRS.3.PL the(M.PL)
pomodor-i
tomato-M.PL
"I like tomatoes."

1.5. ASSIGNING THEMATIC ROLES

Italian disambiguates thematic roles through the interaction of morphological agreement (clitic-NP; subject-verb), WO, likelihood of animacy and stress. For the role of clitic-NP agreement in assigning thematic roles, refer back to 1.3.4 (27).

In (46) disambiguation is due to subject-verb agreement. *Le volp-i* "the foxes" agrees with the verb (PRS.3.PL), whereas *la gallina* "the chicken" does not. *Le volpi* is the subject, *la gallina* is the object (VOS).

46. *Rincorrono la gallina, le volpi*

rincorr-ono	*l-a*	*gallin-a,*	*l-e*	*volp-i*
chase-PRS.3.PL	the-F.SG	chicken-F.SG,	the-F.PL	fox(F.)-PL

"The foxes chase the chicken."

In (47), thematic role assignment is due to the interaction of stress and what Bates, MacWhinney, Caselli, Devescovi, Natale and Venza (1984) call *likelihood of animacy*. Stress singles out the object from the rest of the sentence (it is a case of focalisation). However, morphology is not a cue and focalisation is still ambiguous, as it could be a case of SVO (see 33). Likelihood of animacy solves the puzzle; it is impossible that a house would clean a man. It must be the other way around (hence OVS).

47. *LA CASA, ha pulito Marco*

L-A	*CAS-A,*	*ha*
the-F.SG	house-F.SG,	has

pul-i-to	*Marco*
clean-TV(conj.III)-PTCP	Marco

"Marco has cleaned THE HOUSE (not something else)."

1.6. SUMMARY

Italian is a highly inflected language in which bare roots are bound morphemes. Every noun is either masculine or feminine in gender and nouns, articles and adjectives agree in gender and number. The verbal inflectional system is characterised by three inflectional classes (conjugations), which differ in terms of size, productivity and regularity. Conjugation I is the biggest, the most productive and the most regular, whereas Conjugation II is the *locus* of irregularities, as 95% of its verbs are irregular. A specific verb form is obtained by combining either its stem or bare root with various affixes that carry aspect-mood-tense-person-number information.

Italian is a pro-drop language in which discourse pragmatics determines both null subject realisation and the linear order of the major phrasal constituents. Indeed, Italian is a topic-dominant language (Bates & MacWhinney, 1987) whose flexible WO is discourse-driven.

Thematic roles are disambiguated by the interaction of morphological agreement (clitic-NP; subject-verb), phonetic stress, likelihood of animacy and word order. Morphological agreement is by far the strongest cue on which Italian speakers rely to make sense of utterances (Bates & MacWhinney, 1987; D'Amico & Devescovi, 1993).

2.

A USAGE-BASED APPROACH TO LANGUAGE

Usage-based approaches (UBAs) to LA originate within various linguistic theories that fall under the broad umbrella of "Cognitive Linguistics" (see Croft & Cruise, 2004). The following discussion largely summarises and draws on two main theoretical frameworks: Cognitive Grammar (Langacker, 1987, 1991, 2000, 2008 and 2010) and Construction Grammar (Goldberg, 1995, 1999, 2006). Croft and Cruise's (2004) and Dąbrowska's (2004) works are also sources on which I draw. Needless to say, the following account is by no means an exhaustive and complete overview of either Cognitive or Construction Grammar. Rather, this chapter aims to provide some key background assumptions, which are needed for a broader understanding of how usage-based (UB) researchers account for LA.

2.1. LANGUAGE AND GENERAL COGNITIVE PROCESSES

One of the core claims of Cognitive Grammar (CG, henceforth) is that

> "linguistic structure can only be understood and characterized in the context of a broader account of cognitive functioning"
>
> (Langacker, 1987, p. 64).

Language processing, production and acquisition are accounted for in terms of a few general cognitive processes that are applied and adapted to the specific task of verbal communication (Langacker, 2000, 2008).

Association is the ability to associate and connect one experience and/or concept to other concepts/experiences. For example, we are able to associate a red traffic light with the concept of stopping. Similarly, we are capable of associating the sound /kæt/ with the concept of a cat.

Automatisation (or **entrenchment**): through repetition we are able to master complex routines (rolling a cigarette) in a way that they become automatic and require little or no conscious focus. Linguistically, we are able to produce very long and complex words (e.g. *understandably*) without having to think about the complex tongue and lips movements involved.

Composition is the ability to create a composite structure out of component structures. Lego bricks can be used to build a castle and analogously, words (*kick*, *the*, *I*, *ball*) can be strung together to form a sentence (*I kick the ball*).

Categorisation is the human tendency to interpret the unknown and unfamiliar on the basis of what is known and familiar. We are able to categorise a racoon as an animal even if we have never seen a racoon. In order to categorise B as belonging to the same category as A, we need to either already have an overarching category C, which can subsume both A and B, or form such a category so that both A and B can be grouped together. In front of a set of casino tokens of different colours and shapes, we are able to group them on the basis of their colour and abstract away from differences in shape and vice-versa. This ability to create unified groups (categories) by reinforcing commonalities and filtering out differences is named **schematisation. Schematisation** and **categorisation** are two intimately related processes of **abstraction,** in that they both involve apprehending two or more entities as similar by abstracting away from their differences (Langacker, 2000). Broadly speaking, **schematisation** is the ability to create categories. **Categorisation** is the process through which entity B is associated with entity A because the two share some specifications (e.g. they belong to the same conceptual category)

2.2. LINGUISTIC UNITS

CG posits that language knowledge is about mastering a highly interconnected and structured network of conventionalised linguistic units (Langacker, 1987, 2008). These units can be of three types: **phonological**, **semantic** and **symbolic**, and they are part of networks in which they are connected by **relationships of categorisation**, **composition** and **symbolisation** (Langacker, 1987, 2008).

2.2.1. SYMBOLIC UNITS AND SYMBOLISATION

Symbolic relationships yield **symbolic units** by pairing semantic conceptualisations (such as the concept of a dog) to phonological forms (/dɒg/), which are often referred to as semantic pole and phonological pole, respectively. Simply put, **symbolic units** are **pairings of phonological and semantic units,** linked by **symbolisation**. Conventionally (Langacker, 2008; Dąbrowska, 2004), such units are represented as in fig. 2.1a, in which the upper line represents the concept of DOG (semantic pole), whereas the bottom line represents its phonological realisation (phonological pole). Alternatively, the same unit can be represented in written form as [[DOG]/[dɒg]], where [DOG] represents the semantic pole, [dɒg] the phonological pole and the symbolic relation between the two poles is represented by the slash (/).

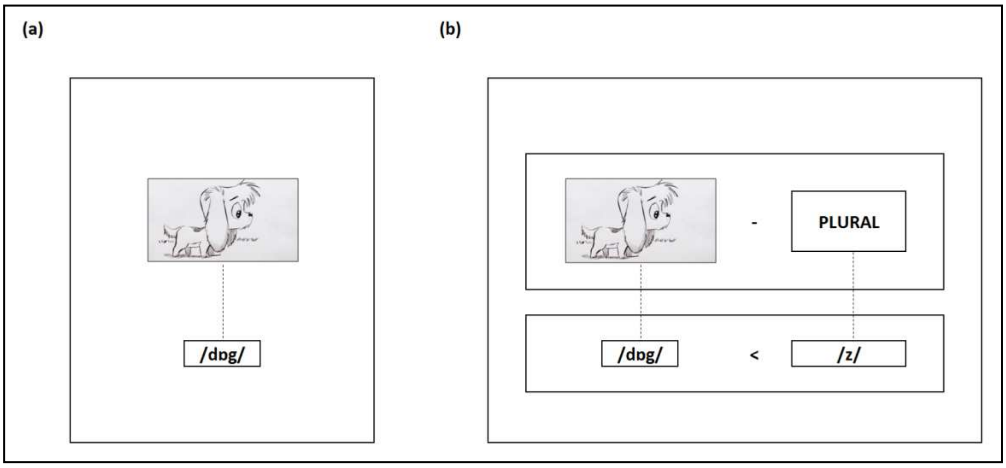


Figure 2.1: Graphic representation of symbolic units. The bottom row represents the phonological pole and the upper row represents the semantic pole. Dotted lines represent relationships of symbolisation between the two poles. The hyphen (-) stands for semantic integration and the symbol "<" stands for linear order (or temporal sequence).

Symbolic units can be either simple or complex. The latter are symbolic units that can be broken down (or analysed) into two or more simpler units (morphologically complex words, phrases and even sentences). *Dogs* is a complex symbolic unit made out of the simple symbolic units [[DOG]/[dɒg]] and [[PLURAL]/[z]], and can be represented in written form as [[DOG]-[PLURAL]]/[[dɒg]<[z]], where "<" represents linear order (or sequential timing) at the phonological pole and the hyphen (-) represents semantic integration (fig. 2.1b). Similarly, *I eat soup* is a pairing of sounds linked to a meaning and can be analysed into smaller units: [[[SPEAKER.DEICTIC]-[EAT]-[SOUP]]/[[aɪ]<[i:t]<[su:p]]].

2.2.2. COMPOSITION

If symbolisation is a relationship that connects phonological and semantic units, both **composition** (or **integration**) and **categorisation** are relationships found between all types of units: phonological, semantic and symbolic.

Composition is the relationship between a composite structure (*dogs*) and its component units (*dog* and *–s*) and can be found at any level of linguistic organisation. At the phonological level, [dɒg] combines with [z] to yield [dɒgz]. At the semantic level, [DOG] combines with [PLURAL] to form the unit [[DOG]-[PLURAL]]. Finally, the symbolic unit [[DOG]/[dɒg]] combines with [[PLURAL]/[z]] yielding the composite symbolic unit [[[DOG]-[PLURAL]]/[[dɒg]<[z]]]. A composite structure (*dogs*) can also serve as a component structure of another, more complex composite structure (*it's raining cats and dogs)*.

2.2.3. CATEGORISATION

Categorisation is the ability to recognise an entity B as belonging to the same category as another entity A: a *sanctioning structure A* (SS) is recognised in a *target structure B* (TS). Categorisation depends on our ability to perceive (and conceive) **schematic relations** amongst entities and conceptualisations (Langacker, 1987, 2000).

Schematic relationships can be thought of as taxonomic hierarchies; the same entity can be conceptualised at different levels of granularity or specificity. In terms of semantic units, we can think of our friend's dog as YURI, LABRADOR, DOG, QUADRUPED, ANIMAL, BEING (Langacker, 2000; Dąbrowska, 2004). In this taxonomic hierarchy, the superordinate element is a schema and the subordinate one is the schema's instantiation. Hence, DOG is both an instantiation of QUADRUPED and schematic with respect to LABRADOR. The relationship between QUADRUPED and DOG is of **full-schematicity** or **instantiation**: the specifications of QUADRUPED are fully recognisable in, and compatible with, the specifications of DOG; the latter specifies the former in more fine-grained detail (DOG is a specific case of QUADRUPED). This is conventionally represented with a solid arrow from the schema to its instantiation (SCHEMA → instantiation;
QUADRUPED → DOG). Because instantiations elaborate schemas, the relationship between the two is also called **elaboration**. Throughout this work, the terms **full-schematicity**, **instantiation** and **elaboration** will be used interchangeably. Relationships of instantiation can be found at any level of linguistic organisation. Phonologically, /dɒg/ is an instantiation of the pattern [[+CORONAL]<[+VOCALIC]<[+DORSAL]], which in turn is an instantiation of the pattern CVC. Symbolically, [[[TALL]-[BOY]]/[[tɑːl]<[bɔɪ]]] is an

instantiation of the pattern *ADJ NOUN* and *I kick you* collocates along the taxonomic hierarchy

I kick you → AGENT-PROCESS-PATIENT → SVO → NVN.

The relationship between subordinate or target structure and sanctioning or superordinate structure (the schema) however, might not be as straightforward and unproblematic. When the specifications of a target or subordinate structure B (TS) conflict with the specifications of a categorising or superordinate structure A (SS), the relationship is one of **extension** (or **partial schematicity**) (Langacker, 1987, 2000, 2008). In order to recognise B as belonging to the same category as A, one must override some of the specifications of the latter, which are not fully met by the former. A typical example of relationships of extension is the different meanings of a polysemous lexical item (Dąbrowska, 2004; Langacker, 2008). Following Langacker (2008), let us assume that a schematic meaning of *ring* is [CIRCULAR OBJECT]. Nevertheless, *ring* can also be used for [CIRCULAR ARENA]. This is possible because some of the specifications of [CIRCULAR OBJECT] (namely the fact that it is an object) are overridden and the term can be extended to describe circular locations. Once unit B (CIRCULAR ARENA) has been categorised as belonging to category A (*ring*), it can serve as the basis for comparison in order to assimilate another entity/conceptualisation into the category. For instance, if a circular arena can be called *ring*, the term can be used to describe a rectangular arena for boxing matches (see Langacker, 2008). Extension is conventionally represented with a dashed arrow

(SS ---► TS; CIRCULAR OBJECT --► CIRCULAR ARENA).

In cases involving extension, category membership is defined in terms of prototypical instance. A **prototype** is the entity/conceptualisation that is more salient/relevant to the category and hence more representative of it. Category membership is then a matter of degree, without any clear cut-off point: the more similar to the prototype a conceptualisation/entity is perceived to be, the more likely it is to be apprehended as belonging to the prototype's category (Langacker, 1987, 2008).

2.2.4. STATUS OF UNITS

Thus far it has been stated that linguistic units are of three types (phonological, semantic and symbolic) and that those units are connected by relationships of symbolisation, composition and categorisation. This raises the question as to what exactly determines which linguistic expressions can be assigned status of unit.

2.2.4.1. Entrenchment

Previously, it has been stated that automatisation is one of the general cognitive processes that are adapted to and for verbal communication. **Entrenchment** is the automatisation of specific linguistic patterns (lexical items, phrases, sentences). The "status of unit" of a linguistic expression is a function of its cognitive salience and entrenchment: the more an expression occurs, the more entrenched it will be in speakers' minds. Importantly, since entrenchment is a function of frequency, it is a matter of degree. Consequently, the "status of unit" of an expression or pattern is also a matter of degree. The result is that linguistic units collocate themselves on a continuum of entrenchment. The more an expression is used, the more it will be cognitively available to speakers and hence easier to retrieve. Conversely, the less it is used, the more difficult it will be to retrieve. Hence, if a symbolic unit, be it simple (*banana*, *cuddle*) or complex (*bananas*, *cuddles*, *unbelievable, the more the merrier, give me that*) occurs with enough frequency and in contexts that are salient enough, it will become a *linguistic gestalt*, easily retrievable as a whole, without the need to "assemble" it (Langacker, 1987, 2000, 2008).

When a particular expression or pattern becomes entrenched in the minds of speakers of a given speech community who use and share it, such an expression/pattern becomes a conventionalised symbol (or unit) of that language.

Speakers can extrapolate those units by:

a) singling them out of the concrete expressions that occur in language use: [[BANANA]/[bəˈnɑː.nə]] is encountered within actual usages of language (*I like bananas*),

b) abstracting schemas which represent schematisations/abstractions from concrete expressions.

(Langacker, 2008)

2.2.4.2. Schematic symbolic units

As for point (b) above, it is worth pointing out how, to some extent, everything emerges from schematisation. Even lexical items that are not particularly polysemous (e.g. [[CAT]/[kæt]]) are acquired through schematisation. A child learning the symbolic unit [[CAT]/[kæt]] might encounter the word when a black cat is present. The second time the cat might be bigger and white and a third one very small and brown. In order to arrive at the conceptualisation [CAT] constituting the semantic pole of the unit, s/he has to abstract away from those differences in colour and size (amongst others) (Langacker, 2000, 2008).

Importantly, a more schematic, complex form-meaning pairing, which stems from generalisations inferred from more specific expressions, can become entrenched and reach status of unit. For example, patterns like *give me that*, *give me the money*, *give me my hat* might trigger a schematisation yielding the unit *give-me-THING_GIVEN* (fig. 2.2) or, more precisely, [[[GIVE]-[ME]-[THING_GIVEN]/[[giv]<[mi:]<[...]]]; (fig. 2.3).

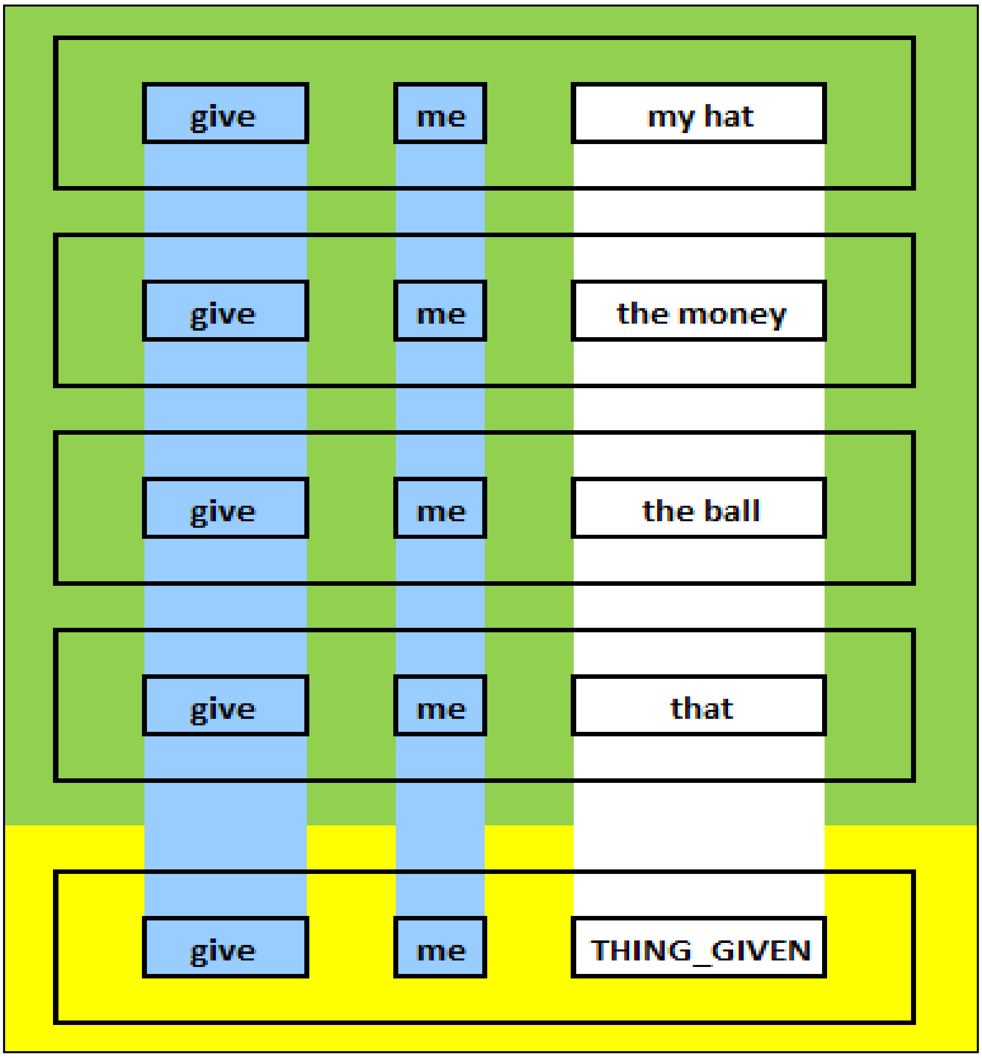


Figure 2.2: abstracting the schematic unit *give-me-THING_GIVEN* (yellow strip) from concrete expressions (green strip). The component parts shared by the schema (yellow strip) and its instantiations (green strip) are highlighted in blue. The variable elements across which *THING_GIVEN* represents a schematisation are highlighted in white.

Give-me-THING_GIVEN is a schema, in that it contains a subpart (*THING_GIVEN*) which is schematic with respect to its instantiations *that*, *the money*, *the ball* and *my hat*. The schema has two main subparts:

a) The subpart *give-me*, whose semantic pole is fully specified at the phonological pole and is recurrent in the schema's instantiations.

b) The subpart *THING_GIVEN* is instead schematic with respect to its instantiations. Its semantic pole symbolises some entity that can be given and its phonological pole is unspecified (represented as [..]), meaning that no phonological restrictions are placed on the elements that can instantiate it (fig. 2.3).

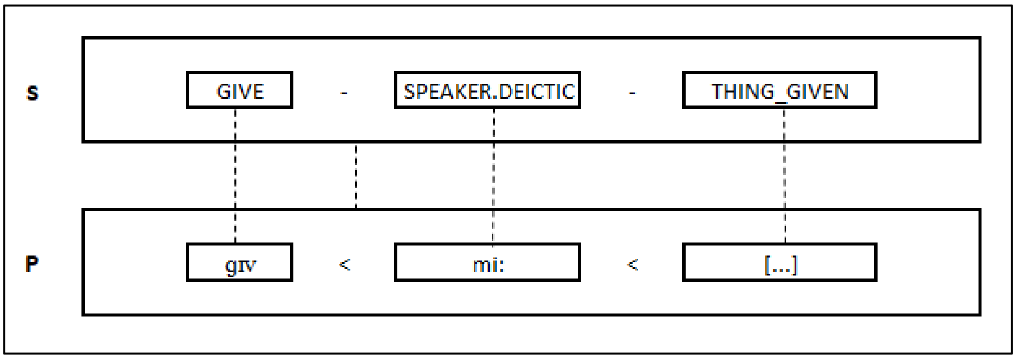


Figure 2.3: the schematic unit *give-me-THING_GIVEN*. The top line represents the semantic pole (S), where hyphens (-) indicate semantic integration. The bottom line represents the phonological pole (P), where the symbol "<" represents linear order.

Conventionally, such schemas are represented with *small letters* indicating elements that are specified at the phonological pole, whereas *CAPITAL LETTERS* represent semantic generalisations that are not specified at the phonological pole (*give-me-THING_GIVEN*). This schema is a formula on which speakers can rely in order to request something.

These types of units are often called **lexically-bound** or **lexically-specific schemas**, because the schematic subpart is bound to, or built around, a subpart which is phonologically fully-specified.

2.2.5. MORE ON SYMBOLIC UNITS

Symbolic units can be simple (*dog*) or complex (*dogs*), the latter being those units that can be analysed into simpler symbolic units. In the previous subsection, I have also implicitly introduced the idea that symbolic units can have different degrees of schematicity, as they can take the form of **schemas** (*give-me-THING_GIVEN*) abstracted by drawing generalisations from other units or expressions.

Schemas are symbolic units that are **not** fully specified at the phonological pole. The phonologically unspecified elements are called *SLOTS* (*THING_GIVEN* in *give-me-THING_GIVEN* is a *SLOT*) and represent generalisations of any sort (morphological, semantic, etc.) across the concrete instantiations from which schemas are inferred.

Symbolic Units fall along a continuum from fully concrete (each element at the semantic pole is phonologically specified, *give me that*) to fully schematic (none of the elements at the semantic pole are phonologically specified, *TRANSFER-RECIPIENT-THING*). Nevertheless, they all are form-meaning pairings. Table 2.1 shows that symbolic units can be of any size and have any degree of specificity.

Table 2.1: Symbolic Units. CAPITAL LETTERS indicate slots, that is, generalisations that represent phonologically-unspecified elements of a (schematic) symbolic unit. *Small letters* indicate elements whose semantic pole is fully specified at the phonological pole (refer back to fig. 2.3). The unit [b…..] indicates words whose initial morpheme is /b/. The psychological reality of such a unit is confirmed by the fact that we can engage in games in which we think of words whose initial phoneme is /b/. Such a unit is unspecified at the semantic pole. Yet, it is partially schematic because part of its phonological pole is (partially) specified.

SYMBOLIC UNITS			
	simple	complex	
fully concrete	banana	kicked	I make those
partially schematic	b……….	PROCESS-ed	MAKER make THING
fully schematic	THING	PROCESS-INFLECTION	AGENT-PROCESS-PATIENT

2.2.6. CONSTRUCTIONAL SCHEMAS

Grammar is often described as a (limited) set of general and productive patterns for assembling complex expressions. Speaking a language undoubtedly implies mastering those patterns in order to produce and understand the many novel expressions speakers encounter on a daily basis (Langacker, 2008). Traditionally, this ability is seen as knowledge of a fairly limited set of (morpho-syntactic) rules and restrictions on them, which determine what is and what is not possible in a given language.

2.2.6.1. Grammatical patterns are symbolic units

CG takes a different approach and claims that language is symbolic in nature and that grammatical knowledge is about mastering a highly interconnected network of symbolic units of different sizes and degrees of specificity (Langacker, 1987).

In CG, regular and productive patterns that are often referred to as rules are captured by **constructional schemas**, i.e. complex symbolic units that are fully schematic (Langacker, 2008). Constructional schemas are form-meaning pairings in which the semantic pole is specified in very schematic terms (AGENT, RECEIVER, THING, PATIENT, PROCESS), whereas the phonological pole is fully unspecified. Fig. 2.4 shows four constructional schemas.

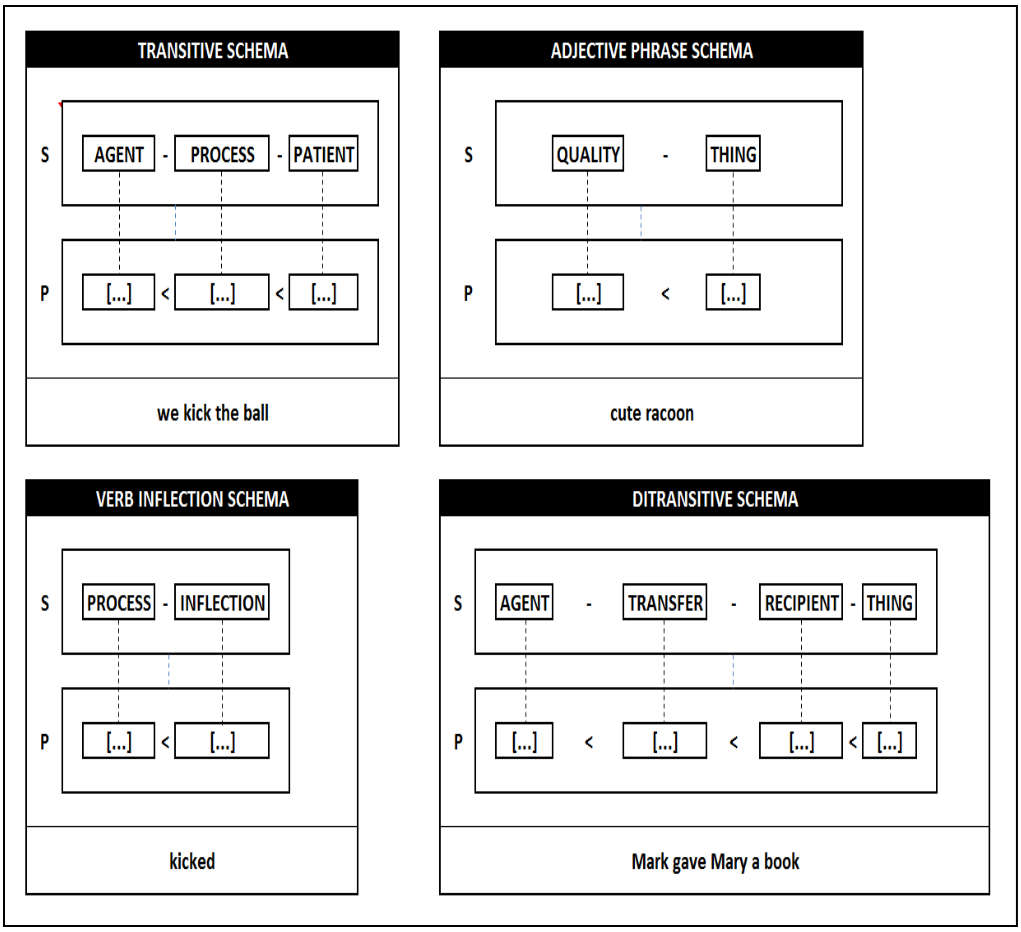


Figure 2.4: constructional schemas. For each schema, its semantic pole (S), its phonological pole (P) and a concrete instantiation of it are provided. Dashed lines indicate relationships of symbolisation. The symbol "<" indicates linear sequence at P. The hyphen (–) indicates integration at S.

Constructional schemas are abstracted from concrete expressions and capture the higher-order regularities that are attested in language use. Hence, the *AGENT-PROCESS-PATIENT* schema is inferred by drawing analogies across the concrete strings that instantiate it (*I kick you*, *you eat that* and so forth). Importantly, constructional schemas and the strings that instantiate them, by definition share the same structure (e.g. the entity acted upon takes post-verbal position in both *AGENT-PROCESS-PATIENT* and *I kick you*). They do however differ in degree of specificity, in that instantiations are more specific at both the phonological and semantic poles. Each pole (semantic and phonological) of instantiation and schema is linked by a relationship of elaboration, as the instantiation specifies the schema in more fine-grained detail (fig. 2.5). Crucially, thanks to such structural and semantic similarities, they are learnable on the basis of language use through processes of abstraction and schematisation (see fig. 2.6).

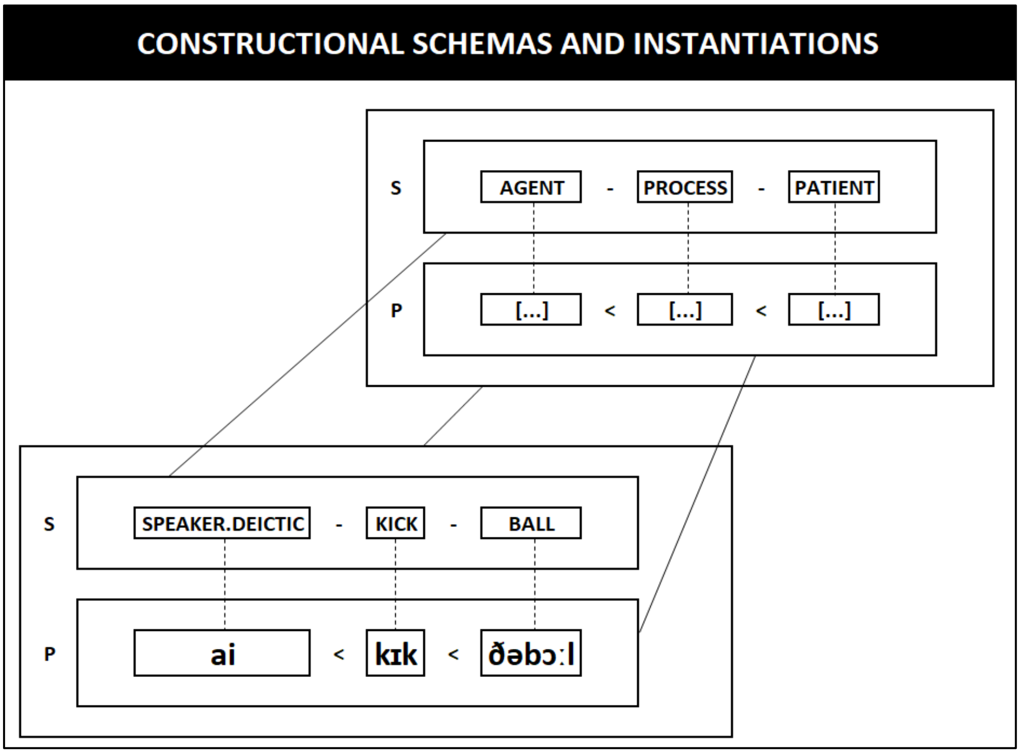


Figure 2.5 The semantic (S) and phonological (P) poles of *AGENT-PROCESS-PATIENT* and *I kick the ball*. Dashed lines represent relationships of symbolisation. Continuous lines indicate relationships of elaboration/instantiation.

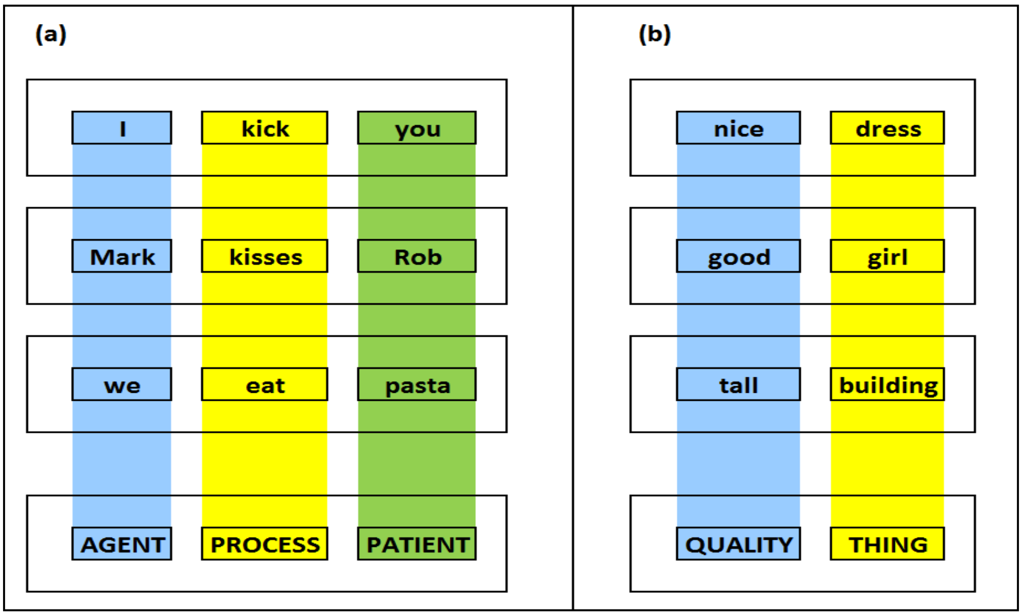


Figure 2.6: Abstracting constructional schemas from their concrete instantiations.

Finally, constructional schemas do not exist per se, as independent entities (or abstract rules), but they are the product of speakers' generalisations of the concrete language they experience. Rather than being independent, separate entities on the basis of which complex expressions are computed, they are immanent in those expressions that instantiate them (Langacker, 2000).

2.2.6.2. Schemas are meaningful constructional patterns

Constructional schemas are meaningful patterns, which substantially contribute to the semantic interpretation of the roles played by the various elements of an utterance. They further specify the way(s) in which those elements are to be integrated together (e.g. subject before the verb) and which kind of "scene" the whole composite structure maps onto (Langacker, 2008).

Goldberg (2006) notes how the semantic interpretation of an expression is the by-product of the semantic integration of:

a) The semantics conveyed by the constructional schema(s)

b) The meaning(s) of the verb(s) involved

c) The meanings of the various arguments involved

d) The pragmatic and discourse context

Thus, (1) is understood as involving a scene in which Mark kicks a ball toward Rob, causing the object to enter into Rob's possession.

1. Mark kicked Rob the ball.

Such an interpretation is possible because (1) is apprehended as an instantiation of the ditransitive construction (*AGENT-TRANFER-RECIPIENT-THING*), which specifies that a preverbal *NML* (*Mark*) causes the first post-verbal *NML* (*Rob*) to receive the second post-verbal *NML* (*the ball*). How the ball is made available to *Rob* by *Mark* is specified by the verb *(kick*) (see Goldberg, 1995, 2006).

2.2.7. MORPHOLOGICALLY COMPLEX WORDS AND BOUND MORPHEMES

Symbolic units – form-meaning pairings - are always the product of some kind of generalisation, of either narrow scope (as the meaning of *cat*, abstracted from different cat types) or wide scope (the *AGENT-PROCESS-PATIENT* schema). Complex symbolic units can be broken down into simpler symbolic units and, because of their composite nature, can also be called **constructions** (Langacker, 2008).

Morphologically complex words are constructions at the word level, whereas syntactic patterns are constructions at the clause or phrase level (Booij, 2010). What follows is that both patterns of morphological integration (e.g. *PROCESS-INFLECTION*) and syntactic patterns (e.g. *TRANSFER-RECIPIENT-THING*) are constructional schemas.

Furthermore, bound morphemes never occur in isolation and their salience to speakers depends on their ability to abstract them from the concrete morphologically complex words in which they appear. **Bound morphemes** can be thought of as **lexically-bound schemas at the word-level**, in which the concrete words or roots with which they combine are specified only schematically. They then have a semantic pole that is only partially phonologically specified (see Booij, 2010). Hence, the past tense morpheme -*ed* and the plural –*s* can be represented as in fig. 2.7 (I ignore allomorphic variation and report the written form).

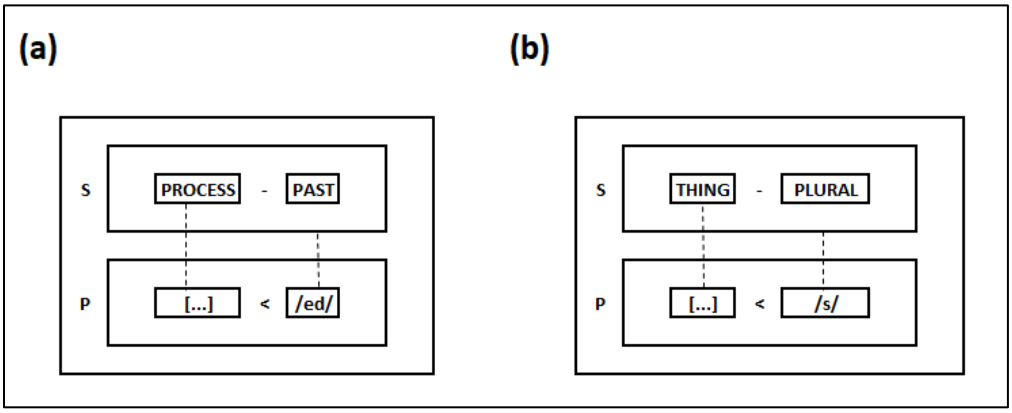


Figure 2.7: bound morphemes as lexically-bound schemas. The upper line symbolises the semantic pole (S), whereas the phonological pole (P) is reported on the bottom line. Dashed lines represent relationships of symbolisation between the two poles. Hyphens represent integration at the semantic pole, whereas the symbol "<" represents temporal sequence at the phonological pole.

2.3. A DYNAMIC AND HIGHLY INTERCONNECTED NETWORK

The main claim of both Cognitive Grammar and Construction Grammar is that language competence is about mastering an enormous, highly redundant and highly structured inventory of symbolic units. These symbolic units can be of any size, can have any degree of schematicity and are linked to each other by relationships of categorisation (partial and full schematicity) and composition.

This inventory takes the form of a highly interconnected **network** and corresponds (or better, it also includes) what is traditionally referred to as grammar in the form of constructional schemas.

2.3.1. THE NATURE OF THE NETWORK

Language can be thought of as a **space** of constructional possibilities that is occupied by a network of symbolic units, which are related (or linked) by relationships of categorisation and composition (or integration). Figure 2.8 is one possible representation of a (very) small part of such an inventory and takes as examples Langacker's (2000, 2008) graphs[15].

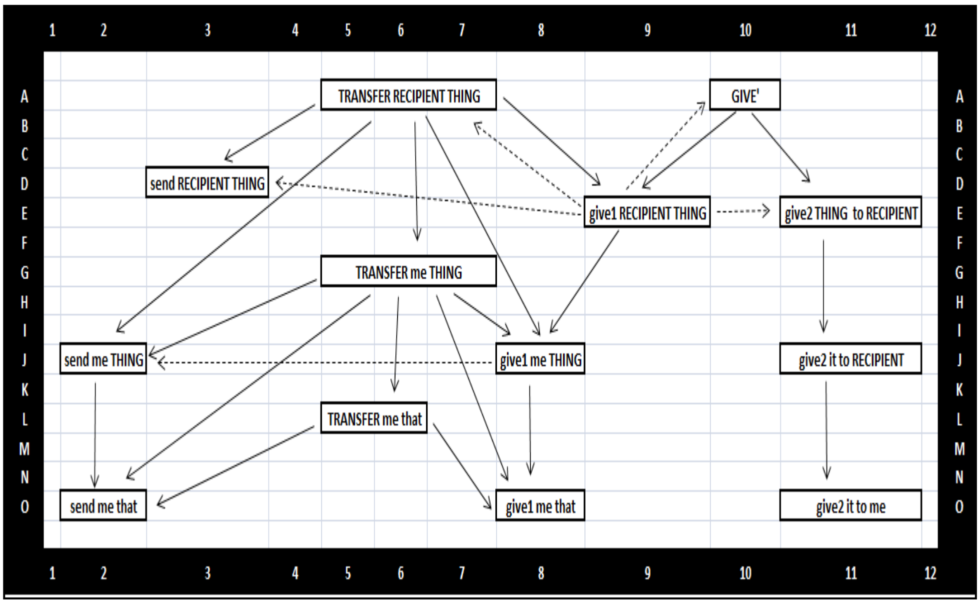


Figure 2.8: network of symbolic units (based on Langacker, 2000, 2008). Symbolic units are enclosed in rectangles. *Small letters* indicate elements specified at the phonological pole. *CAPITAL LETTERS* indicate slots. Dashed arrows indicate relationships of extension. Solid arrows indicate relationships of elaboration.

Expressions of any size and degree of schematicity, if recurrent enough, can become entrenched and acquire status of unit. A unit is a pre-packed linguistic *gestalt* that does not need to be assembled, as it is available and retrievable as a whole.

[15]Note, however, that it is not claimed that knowledge of the units illustrated in fig. 2.8 can be precisely captured by such a graphic representation. Rather, the figure is meant to have explanatory value.

Thus, if *give me that* (fig. 2.8, O8) and *give it to me*[16] (fig. 2.8, O11) occur with enough frequency, they can become entrenched and reach status of unit. In this case, they would be units which are both fully concrete (each element of the semantic pole is specified at the phonological pole) and complex (they are analysable into component parts). Since *give me that* (fig. 2.8, O8) is a complex unit, the units out of which it is composed (*give*, *me* and *that*) are units themselves, which are connected by relationships of composition (or integration).

At the same time, many expressions such as *give me my money*, *give me my hat*, *give me the ball* might as well occur with salient type and token frequency. Speakers would then be able to extrapolate the symbolic unit *give-me-THING* (fig 2.8, J8), which is a lexically-bound schema on which speakers can rely in order to request something (fig. 2.2 and 2.3).

Similarly, if speakers witness many utterances like *give mum her keys*, *give him some cake*, *give us our books*, the schema *give-RECIPIENT-THING* (fig. 2.8, E9) might become an entrenched and cognitively salient template one can rely on in order to express actions of giving.

Give-RECIPIENT-THING (fig. 2.8, E9) and *give-THING-to-RECIPIENT* (fig. 2.8, E11) are linked by a relationship of extension (dashed arrow). The specifications of the two constructions differ in many respects, such as information structure and the kinds of elements that can instantiate THING and RECIPIENT (Goldberg, 1995). Nevertheless, they both express the act of giving THING to RECIPIENT; such a shared meaning is represented by the schematic and decontextualised *GIVE'* (fig. 2.8, A10).

Finally, both *give-RECIPIENT-THING* (fig. 2.8, E9) and *send-RECIPIENT-THING* (fig. 2.8, D3) imply the transfer of THING to RECIPIENT. This shared property is embodied by the constructional schema *TRANSFER-RECIPIENT-THING* (A5-A7), which they instantiate by specifying it in more fine-grained detail. *TRANSFER-RECIPIENT-THING* may also be

---

[16] *Give1* and *give2* in the figure are meant to differentiate instances of the verb *give* when it appears in different constructions. The various meanings of a lexical item are acquired and establish themselves because speakers abstract away from the contexts in which they appear, retaining their common features (see also *2.2.4.2* and *2.3.2.1*)

instantiated by (and inferred from) other expressions such as *email them your complaints* or *post me everything I need to know*.

Importantly, *TRANFER-RECIPIENT-THING, give-RECIPIENT-THING, give-me-THING* and *give-me-that* hold relationships of instantiation (solid arrows) and are organised taxonomically:

TRANFER-RECIPIENT-THING ⟶ *give-RECIPIENT-THING* ⟶ *give-me-THING* ⟶ *give-me-that*.

2.3.2. PARTITIONING THE SPACE

2.3.2.1. Complex categories

The huge language network is analysable into smaller sub-networks, in which families of formally and semantically related units group together and form **complex categories**, each of which occupies a certain space of **semantic and constructional** (**formal**) possibilities (Langacker, 2008). Complex Categories are the bedrock of virtually any aspect of linguistic organisation: different meanings of a lexical item, allomorphic relationships, morphological suppletion and families of grammatical constructions, all represent complex categories (Langacker, 2008). Networks of relationships between constructional schemas and their instantiations also represent **complex categories**. For instance, the network of symbolic units that constitutes knowledge of the ditransitive construction occupies a **semantic and formal** space, albeit a limited one. In such a space, expressions mapping onto life events in which *THING* is made available to *RECIPIENT* (in either literal or metaphorical sense) group together. Such expressions are also formally related in that they all instantiate the pattern *PROCESS-NML1-NML2* (or *TRANSFER-RECIPIENT-THING*). In fig. 2.8, columns 1-8, one part of the space occupied by the ditransitive construction is represented.

The different meanings and structural frames of lexical items also constitute complex categories: part of the complex category representing the verb *give* is represented in figure 2.8, columns 8-12. In 2.2.4.2, it has been pointed out that even lexical items that are not particularly polysemous are the by-products of some kind of generalisation. The psychological reality of GIVE' (fig. 2.8, A10) is the by-product of speakers abstracting this decontextualised meaning by filtering out (or abstracting away from) the different structural frames in which the verb is attested. However, a verb does not actually exist outside its structural frames, as the latter are inherent in its very conceptualisations and meanings and, as such, they are essential aspects of its characterisations. Structural frames are lexically-bound schemas that schematically specify the arguments taken by a verb and represent actual mini-grammars (Langacker, 2000, 2008; Dąbrowska, 2004). The complex category that represents knowledge of the verb *give* is "composed" by:

a) the network of schemas specifying the frames in which it appears
b) the morphological forms it takes (*give*, *gives*, *given*, *giving*)
c) its decontextualised schematisation *GIVE'*.

Knowledge of *give* is about mastering the whole network (the whole complex category) and the way it occupies the linguistic space.

Finally and crucially, a complex category may "share" some of its units with other complex categories. In a sense, it could be said that complex categories are linked to each other by "**shared nodes"**: in fig. 2.8 *give-RECIPIENT-THING* (E9) is part of the network describing the meanings and distributional properties of *give* (columns 8-12) **and** it is also part of the network constituting knowledge of the ditransitive construction (columns 1-8), in that it instantiates both *TRANSFER-RECIPIENT-THING* (A5-A7) and *GIVE'*(A10). Hence, *give-RECIPIENT-THING* (fig. 2.8, E9) and its instantiations *give-me-THING* (fig. 2.8, J8) and *give-me-that* (fig. 2.8, O8) represent the formal (or constructional) and semantic space shared by the complex category of the lexical item *give* and the complex category of the ditransitive construction. "Shared nodes" yield quite substantial overlaps, making it possible to describe the whole space as a gigantic network

2.3.2.2. Redundancy and low-level schemas

Networks are said to be highly **redundant,** in that the same information can be stored at different levels of specificity. A pattern that could be represented only once at the most schematic level (*TRANSFER-RECIPIENT-THING*) may as well be represented in the system by different units that instantiate that very same pattern in more specific terms (*give-RECIPIENT-THING*, *give-me-THING*), as long as those more specific expressions become entrenched enough to attain status of unit.

Redundant information in a network representing the structural and semantic properties of a syntactic pattern (transitive and ditransitive construction, for instance) often coincides with what is traditionally called the argument structure of a verb. In fig. 2.8, *TRANSFER-RECIPIENT-THING* (A5-A7) is also represented at a more specific level by *give-RECIPIENT-THING* (E9) and *send-RECIPIENT-THING* (D3). These more specific units are what Langacker (2000, 2008) calls **lower-level schemas**; i.e., schemas that, like constructional schemas, are abstracted by generalising from specific expressions, but represent generalisations of a narrower scope than the constructional schemas they instantiate. **Lower-level schemas** are essential to linguistic organisation and, in many respects, they are more salient and more viable than constructional schemas (Langacker, 1987, 2000).

If complex categories can be metaphorically described as families of formally and semantically related symbolic units which occupy a given space of constructional possibilities, then languages can be said to rarely exploit all potential structural options. Indeed, permitted patterns often cluster only in certain areas (Langacker, 2000, 2008). The precise information on how a complex category occupies a given structural and semantic space is provided by the **whole** network of symbolic units (the complex category as a whole). Indeed, more often than not, fully general constructional schemas might not be available or, if potentially available, might not be entrenched nor salient enough to be evoked in language production and processing (Langacker, 2008).

For instance, Italian allows both *OV* and *VO* linear order (see (2) and (3) below), as long as certain pragmatic and lexical conditions are satisfied.

2. *Ho letto un libro*
ho *l-e-tto*
have(PRS.1.SG) read-TV(conj.II)-PTCP
un *libr-o*
a(M.SG) book-M.SG
"(I) have read a book."

3. *UN LIBRO, ho letto*
UN *LIBR-O,* *ho* *l-e-tto*
a(M.SG) book-M.SG have(PRS.1.SG) read-TV(conj.II)-PTCP
"A BOOK, (I) have read."

This flexibility is allowed with object NPs and strong pronouns. Hence, Italian speakers might well infer the schema [O+V], where the plus (+) indicates that the phonological pole does not specify the linear order.

However, when a clitic pronoun is involved, the order Vo[17] is obligatory when the verb is [-FINITE] (imperatives, infinites and gerundives). Conversely, when the verb is [+FINITE] the only possible order is oV. Hence, if Italian speakers solely relied on the higher-level schema [O+V], they might produce ungrammatical sentences such as (4), instead of the grammatical (5).

4. **Mangiai lo*
mangi-a-i *l-o*
eat-TV(conj.I)-PST.1.SG clitic.3.ACC-M.SG
"I ate it."

5. *Lo mangiai*
l-o *mangi-a-i*
clitic.3.ACC-M.SG eat-TV(conj.I)-PST.1.SG
"I ate it."

The general pattern [O+V] is instantiated by several lower-level schemas which specify the structural possibilities Italian allows when a DO and a verb are combined together (fig. 2.9).

---

[17] *CAPITAL "O"* refers to NP and tonic or strong pronoun DOs; *small "o"* refers to accusative clitic pronouns. See also footnote 11.

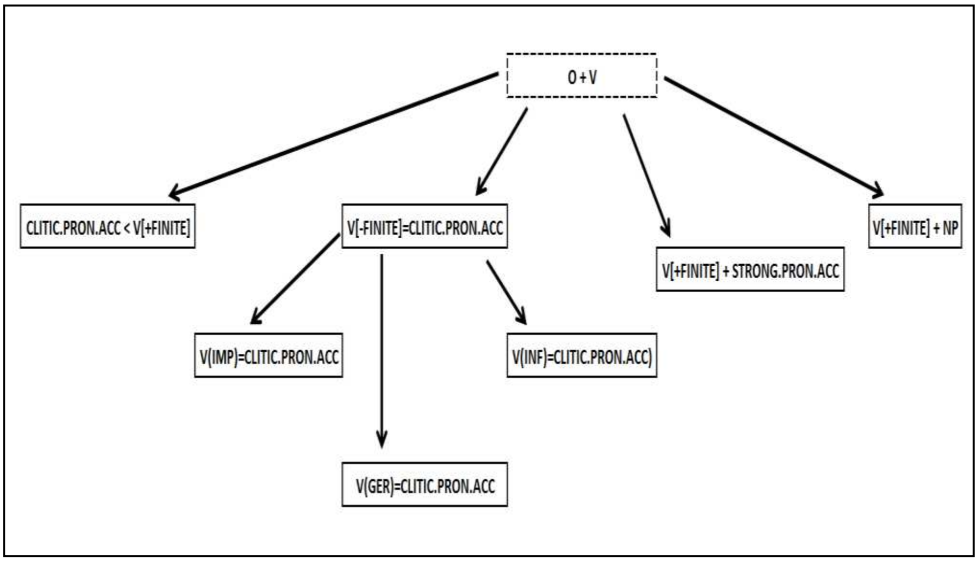


Figure 2.9: Verb and direct object in Italian. The symbol "<" indicates compulsory linear order. The symbol "=" indicates morpho-phonological merging. The symbol "+" indicates that the linear order is free (not specified). The dashed rectangle in which *O+V* is enclosed indicates that the highest-level schema is unlikely to be an entrenched unit available to sanction linguistic expressions. Arrows indicate relationships of elaboration/instantiation.

Knowledge about the possible combinations of verb and DO in Italian is not about mastering the more general schema [O+V]. Rather, it is about mastering the whole network, including low-level schemas and the structural possibilities that these describe.

Thinking of language competence as mastery of the whole network accounts for a phenomenon called **blocking**: a more regular and general pattern does not apply/generalise to certain patterns or words (Langacker, 2008). Since the lower-level schema *CLITIC.PRON.ACC<V[+FINITE]* is an entrenched unit of the network, its presence prevents (or blocks) speakers from uttering sentences like (4).

2.4. LANGUAGE USE

The (main) function of language is communicating one's conceptualisations through sound-meaning pairings. In order to communicate successfully, speakers rely on a vast and highly structured inventory of symbolic units. The task then, is about retrieving the linguistic units that most closely match an intended conceptualisation, that is, it is about retrieving linguistic units (form-meaning pairings) in order to categorise a Usage Event (UE, henceforth). Broadly speaking, a usage event is a specific and unique usage of language that conveys speakers' communicative intentions (see Langacker, 2000, 2008).

Thus, if Mark wanted to order Claire to give him something, the unit *give-me-that* could be retrieved (assuming that such a unit is available as a whole in his inventory). Since the semantic pole of such a unit matches Mark's conceptualisation, he could rely on it for expressing his communicative needs. The semantic pole [[GIVE]-[ME-[THAT]] is linked by symbolisation (/) to its phonological pole [[gɪv]<[me]<[ðæt]]; hence, Mark would know that the appropriate sound for his intention is the string of phonemes /gɪvmeðæt/. Similarly, Claire would recognise /gɪvmeðæt/ as a unit and link it to its semantic pole [[GIVE]-[ME-[THAT]] and would therefore understand his communicative intention.

This is the case when a conceptualisation is fully symbolised by, and included in, a symbolic unit. However, more often than not, symbolic units do not fully match conceptualisations. In this second case, speakers will have to "assemble" a **novel sentence** out of the units available in their own inventories.

UBAs recognise two main processes through which speakers assemble novel utterances: juxtaposition and superimposition (Dąbrowska, 2004, 2014).

2.4.1. JUXTAPOSITION

Juxtaposition is the operation through which a component unit A is attached to either end of another component unit B. Thus, the juxtaposition of *give me that* and *now* could yield either (6) or (7)

6. Give me that, now.
7. Now, give me that.

Rather than purely syntactic, juxtaposition has a paratactic nature, in that, although the meanings of A and B are to be integrated, the grammar does not specify how and posits no restrictions on that (Dąbrowska, 2004).

2.4.2. SUPERIMPOSITION

Superimposition is the operation through which a schematic subpart of a symbolic unit_A (the slot of a schema) is elaborated in more fine-grained detail by another unit_B. Thus, (8) could be assembled by retrieving the lexically-specific schema (9).

8. I need a chair.
9. *NEEDER-need-THING_NEEDED.*

In (9) the arguments of the verb are specified only schematically on the semantic pole and not specified at all at the phonological pole. Hence, the speaker must specify (or elaborate) those schematic arguments, so that both their semantic and phonological forms match her conceptualisation. This is obtained by superimposing *I* and *a chair* over the slots *NEEDER* and *THING_NEEDED*, respectively (fig. 2.10). Phonological and semantic superimposition happen simultaneously (Dąbrowska, 2004).

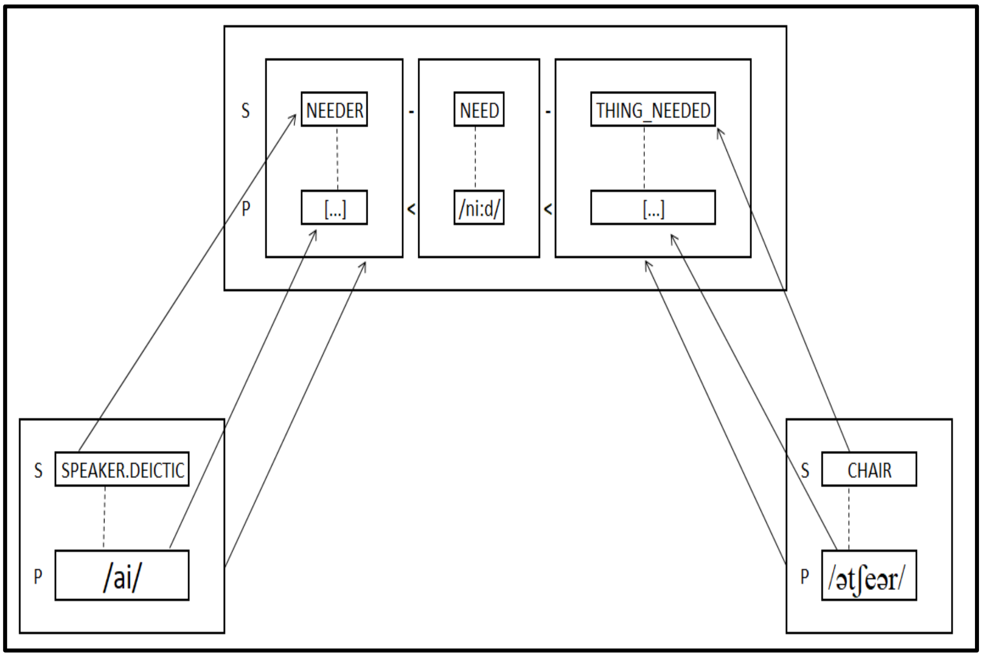


Figure 2.10 Superimposition: dashed lines indicate relationships of symbolisation between semantic (S) and phonological (P) pole. Arrows indicate that the more specific units elaborate the more schematic ones at semantic, phonological and symbolic levels.

Superimposition applies at the clause level (as above), but also at the word level. For example, *dogs* could be derived by superimposing the units [[DOG]/[dɒg]] over the slot of the schema *THINGs* (or [[[THING]-

[PLURAL]]/[[...]<[z]]]) (fig. 2.11b)[18]. Fig. 2.11a is a more appropriate reformulation of fig. 2.1b, because *dog* is inserted *within* the plural schema.

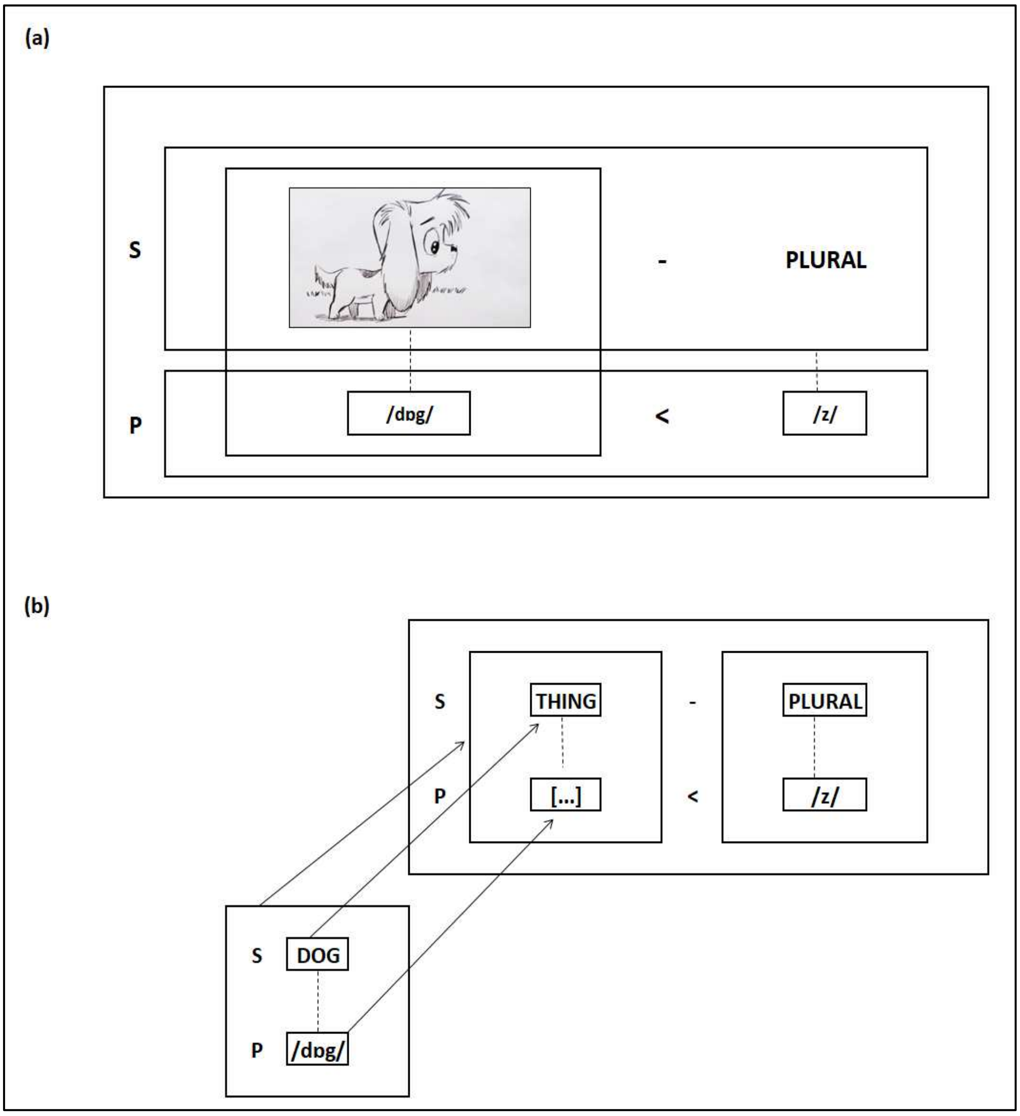


Figure 2.11: *dogs* derived by superimposing *dog* and *THING*s. Arrows indicate relationships of elaboration (from the filler to the schema). Dashed lines represent symbolic relationships between phonological pole (P) and semantic pole (S). The hyphen (-) represents semantic integration and the symbol "<" indicates linear order.

2.4.2.1. Meeting the slot requirements

In fig. 2.10 *a chair* can be superimposed over the slot *THING_NEEDED* because it meets the slot's specifications. A chair is something that can be needed and hence meets the semantic specifications of the slot *THING_NEEDED*. Thus, the relationship between *THING_NEEDED* and *a chair* is of elaboration (or full schematicity) in that *a chair* elaborates in more

[18] Assuming that *dogs* is not an available unit retrievable as a whole, which is unlikely.

fine-grained detail the specifications of the slot (it instantiates the slot)[19]. This relationship of elaboration holds at phonological, semantic and symbolic level. The elaborating unit is often referred to as **filler,** as it is used to "fill the slot".

However, language use is not an exact science and speakers, more often than not, stretch its limits. For instance, *mum* could be used to fill the *PROCESS* slot of *I-don't-know-how-to-PROCESS*, as in (10). When a slot is elaborated by a filler that does not (fully) match its specifications, this is a case of **functional coercion**.

10. I don't know how to mum.

Such usages of language represent one of the main sources of linguistic innovation. Indeed, functional coercion is often the driving force behind young children's creative ungrammatical sentences (see 3.6.2). For instance, (11) could have been assembled by retrieving the constructional schema *ADJ NOUN*, but *ADJ* has been elaborated by *badly*. Being an adverb, *badly* cannot be classified as an instance of *ADJ*; the result is an ungrammatical sentence.

11. **badly boy.*

2.4.2.2. Superimposing two schemas

Sometimes the units superimposed are two lexically-bound schemas: in these cases, the recurrent, phonologically-specified subpart of unit A elaborates the schematic subpart of unit B, and vice versa (Dąbrowska, 2004; Dąbrowska & Lieven, 2005). Hence, (12) could be derived by superimposing (13) and (14).

12. eat them!

13. Eat-THING_EATEN.

14. PROCESS-them.

(13)'s *eat* elaborates (14)'s *PROCESS* and, at the same time, (14)'s *them* elaborates (13)'s *THING_EATEN* (fig 2.12). Henceforth, I shall refer to the superimposition of two schematic units as **mutual superimposition**.

[19] "a chair" elaborates in finer grained detail the semantic concept of *THING* (in Langacker's (2008) terms).

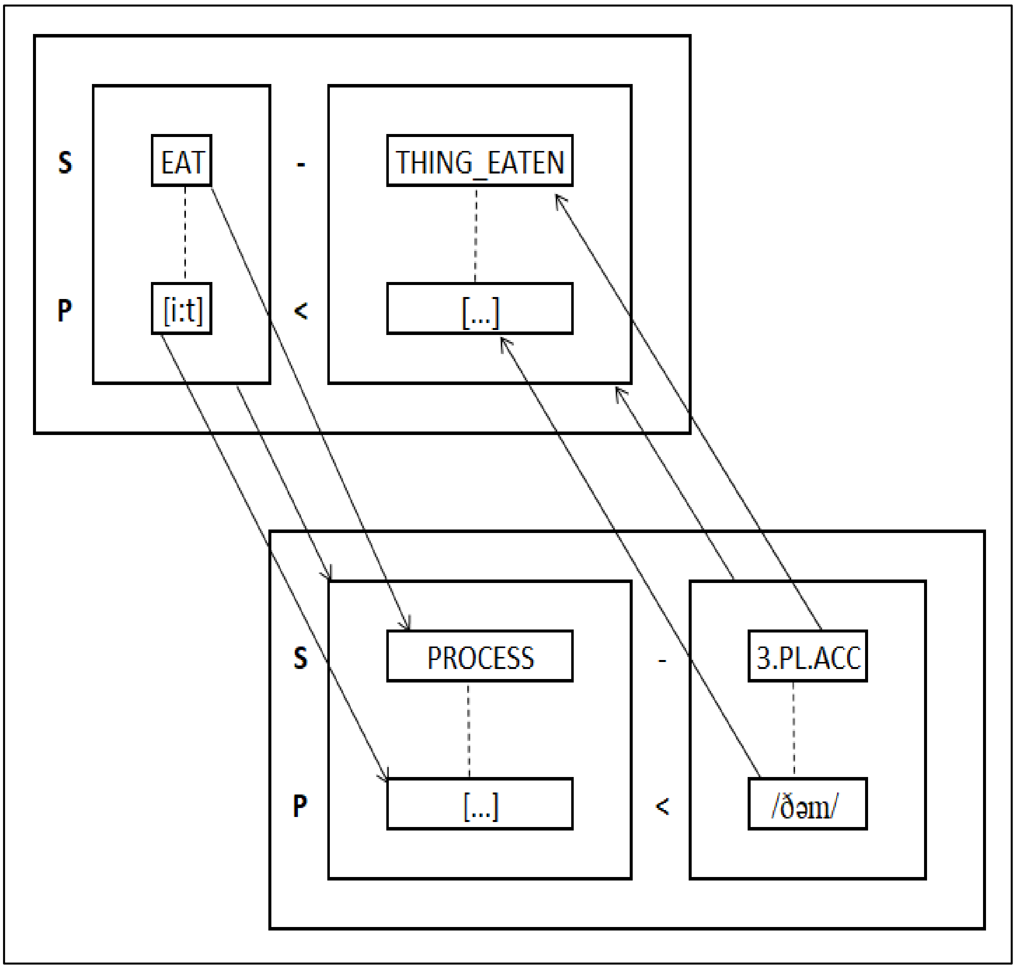


Figure 2.12 Mutual superimposing of two lexically bound schemas. Solid arrows indicate relationships of elaboration (from the more specific unit to the more schematic one). Dashed lines indicate relationships of symbolisation between semantic pole (S) and phonological pole (P). Hyphens (-) represent semantic integration and the symbol "<" indicates linear order.

2.4.2.3. The psychological reality of superimposition

Tomasello (2006b) reports on a study of Brown and Kane (1998) in which two-year-olds were trained to perform certain actions (e.g. pulling) with a specific object (a stick). Participants were then given "transfer problems" to solve, such that the same action could (creatively) be performed with/on different objects. Children carried out the task successfully - i.e. they pulled sticks, ropes, etc. This suggests that that they were capable of forming sensory-motor schemas – in Piaget's (1952) terms - in order to perform the same action with/on a variety of objects (see Tomasello, 2006b).

Such an outcome could be regarded as the non-linguistic counterpart of forming a schema *pull-X* and elaborating the slot with the name of any object that can be pulled (*pull the stick*, *pull the rope*, etc.).

It could therefore be argued that superimposition originates in our ability to form categories (schematisation) in order to infer behavioural patterns (*pull-THING*_PULLED) that contain flexible, yet constrained and coherent parts or slots (e.g. THING_PULLED) to be "played around" with by "filling" them with appropriate material.

2.4.3. SELECTING THE UNITS

At any given time, a particular conceptualisation can potentially be expressed and assembled through different units, each of which is an equally appropriate candidate to categorise the intended conceptualisation. Hence, a set of units that can potentially serve the communicative needs is activated. Langacker (2000, p. 15) calls such a set of units "**Activation Set"**. Whether or not a particular unit matches the communicative intentions of the speaker is a matter of degree. Some units will match those communicative intentions quite closely, while others will match them to a lesser extent and some not at all. Our memory is content-addressable (Dąbrowska, 2004), which means that, if A and B are often associated, retrieving either one will very likely bring about the retrieval of the other (Langacker, 2000). If a speaker wanted to utter something like (15), she might retrieve (16), but because the string *where are you* is likely to be associated with *going* (the association is very strong to my sensitivity, at least), also (17) might be activated. Because of (17)'s activation, even (18) might enter the activation set. Other putatively appropriate units may be (19) and (20). Importantly, in the same way that the extent to which a linguistic unit matches a speaker's communicative needs/intentions is a matter of degree, so is its activation status (see Langacker, 2008). These potential candidates "compete" to categorise the speaker's communicative intentions and, as they do that, they become more or less activated. The process is one in which some activated units will be mutually inhibitory – e.g. (17) and (20) - whereas others will reinforce each other – e.g. (16), (19) and (20) (see Langacker, 2000, 2008). Eventually, some units will be "deactivated", whereas others will become more strongly activated and be selected to assemble the target sentence. The units selected are the **active structures** which categorise the expression (Langacker, 2000).

The selection of activation set and active structures is, however, a matter of milliseconds (Langacker, 2008).

15. Where are you hiding?

16. Where are you PROCESS-ing?

17. Where are you going?

18. Going.

19. WH are you PROCESS-ing?

20. Hiding.

Most likely, the choice of which units are selected will depend on their entrenchment, degree of overlap with the conceptualisation to be communicated and contextual priming (Langacker, 2000, 2008).

2.4.3.1. Entrenchment

Entrenchment is a function of frequency: more frequent units are more entrenched. The more a pattern (be it concrete or schematic) recurs, the easier it will be to activate (Langacker, 2010). Hence, more entrenched units are more likely to be selected to assemble a specific target sentence (Langacker, 2000; Dąbrowska, 2004).

2.4.3.2. Degree of overlap with speaker conceptualisation

All things being equal, more specific units are more likely to be selected because they match speakers' conceptualisations more closely. There are two levels of specificity that are worth exploring: concreteness and complexity (size).

More concrete units (whose semantic pole is fully phonologically-specified) match conceptualisations more closely as they are specified in more fine-grained detail. Thus, if a speaker wanted to say *give me that* and the string were a unit, it would perfectly match her communicative intention. A more schematic unit such as *give-RECIPIENT-THING* would still match the speaker's conceptualisation, but theme and recipient would be specified only schematically and would need to be elaborated. The elaborative distance – in Langacker's (2000) terms (see 2.5.1) - between *give me that* (TS) and *give-*

RECIPIENT-THING (SS) is greater than it is between *give me that* (TS) and *give-me-that* (SS). As a consequence, more specific units are thought to be less computationally demanding, as they potentially allow speakers to "assemble" a target sentence through fewer operations. *Give-me-that* requires no operations, whereas *give-RECIPIENT-THING* requires the speaker to elaborate two slots.

Complexity is also a factor, in that more complex units are also likely to carry greater overlap with speakers' conceptualisations, because they are more specific by nature. For instance, the string *eat them* could be derived from (21) and (22), as in fig. 2.12, or from (21) and (23).

21. Eat-THING_EATEN.
22. PROCESS-them.
23. Them.

The speaker's conceptualisation is about eating a plural entity and it contains both the action of eating and the entity eaten. *Them* (23) perfectly matches the conceptualisation of the entity eaten. However, *PROCESS-them* (22) perfectly matches the conceptualisation of the entity eaten **and** schematically matches the action of eating (the *PROCESS* slot). *PROCESS-them* (22) is as specific as *them* (23) in terms of the entity that undergoes the action. It is also **more** specific than *them* (23) with respect to the action, as the former and not the latter contains an action undergone by that plural entity (even if only in a schematic form).

Dąbrowska (2004) points out that preference for larger and more specific units could account for why speakers normally make very few overgeneralisation and agreement mistakes. A unit such as *PROCESS-them* prevents speakers from filling the slot *THING_EATEN* with *they* (which is 3.PL as well as *them*), yielding the ungrammatical **eat they*.

Importantly, the choice of which units are selected to categorise a speaker's communicative intention is determined by a trade-off between entrenchment and specificity. For instance, in categorising a specific communicative event (*give that old man your new teddy bear*), a less specific but more entrenched unit (*give-RECIPIENT-THING*) might be chosen over a more specific but less entrenched one (*give-that-THING-your-THING*).

2.4.3.3. Conflicting Specifications

It might sometimes be the case that the selected units contrast in some of their specifications; in those cases, speakers must resolve the conflict. A case in point is question formation (see Dąbrowska, 2004, pp. 217-219, for an identical argument with long-distance questions).

Let us assume that a speaker's conceptualisation is (24)

24. What are you kicking?

Kick is a highly transitive verb, which is unlikely to be attested without a DO. Hence, a speaker's representation of *kick* is likely to be schematic (25). Let us assume that the speaker activates (25) and (26)

25. Kick-KICKEE.
26. What-are-you-PROCESSing?

(25) and (26) conflict in that the latter specifies that the DO (*what*) must take initial position, whereas (25) specifies a post-verbal *KICKEE*. In order to assemble the sentence, the speaker must resolve the conflict. Either, (25)'s specifications overrule (26)'s, yielding (27), or vice versa, yielding (28).

27. You are kicking what?
28. What are you kicking?

The above discussion suggests that composite symbolic units can often be potentially derived in alternative ways. This can even be the case with identical component units.

For example, (29) could be assembled by either:

a) juxtaposing (30) and (32), yielding *PROCESS them now*. Subsequently, *PROCESS them now* and *eat-THING_EATEN (31)* are superimposed.

or by

b) superimposing (30) and (31), yielding *eat them*, which is subsequently juxtaposed to *now* (32).

29. Eat them, now!

30. PROCESS-them.

31. Eat-THING_EATEN.

32. Now.

Different types of constituencies are a function of the compositional patterns one can select to assemble an utterance (see Langacker, 2008).

2.5. THE DYNAMICITY OF THE SYSTEM

2.5.1. ELABORATIVE DISTANCE

When they use language, speakers are able to judge whether a particular expression is possible (grammatical) or not. Such judgements are matters of categorisation: a TS (a concrete expression occurring in language use, i.e. in a UE) is judged/categorised in light of an SS (one or more units in the inventory). Categorisation can be thought of as elaborative distance between TS and SS (Langacker, 1987). For instance, speakers judge *give me that* as possible because it is either identical to a concrete unit (*give-me-that*) or an instantiation of a more schematic one (*TRANSFER-RECIPIENT-THING*). The elaborative distance is null in the former case, and it is of full schematicity in the latter.

When the elaborative distance is greater and more problematic, one finds relationships of extension between TS and SS. The next section walks the reader through an example that aims to illustrate the role played by extension, schematisation and entrenchment in the development of complex categories.

2.5.2. THE DITRANSITIVE COMPLEX CATEGORY

Since the “status unit” of an expression is a by-product of its degree of entrenchment and hence dynamic in nature, complex categories (which are networks of symbolic units) are necessarily dynamic too.

According to Goldberg (1995), the ditransitive construction has five senses that can be thought of as extensions of the prototypical meaning *X causes Y to receive Z* (table 2.2).

Table 2.2: the ditransitive construction: prototypical and extended meanings (based on Goldberg, 1995, pp. 38 and 72)

THE DITRANSITIVE CONSTRUCTION				
	PARAPHRASE	types of verbs that can appear in the construction	example	(putative) constructional schema
CENTRAL MEANING (prototype)	X causes Y to receive Z	verbs of giving, continous causation, instantaneous causation and ballistic motion	Joe gave Sally the ball	TRANFER NML NML
EXTENDED MEANING	X causes Y to receive Z, if X complies with what asserted	verbs of giving associated with satisfaction	Joe promised Bob a car	SATISFY NML NML
EXTENDED MEANING	X enables Y to receive Z	verbs of permission	Joe permitted Chris an apple	ENABLE NML NML
EXTENDED MEANING	X causes Y not to receive Z	verbs of refusal	Joe refused Bob a cookie	REFUSE NML NML
EXTENDED MEANING	X intends to cause Y to receive Z	verbs of creation and obtaining	Joe baked Bob a cake	CREATION NML NML
EXTENDED MEANING	X acts to cause Y to receive Z at some future point in time	verbs of future transfer	Joe bequeathed Bob a fortune	FUTURE.TRANSFER NML NML

Complex categories are often built around one or more **prototypes**, from which other nodes radiate. Although prototypes need not be the nodes from which the category develops, for the sake of argumentation, I shall assume that the ditransitive complex category developed from the prototypical *TRANSFER-NML1-NML2*[20]. A hypothetical primitive nucleus of the category is represented in fig. 2.13.

[20] Throughout the following subsection, I adopt the coding used by Langacker (2008) and Goldberg (1995). I shall therefore use more general labels for the arguments associated with verbs and their patterns, such as *TRANSFER-NML1-NML2*, rather than more specific labels such as *TRANSFER-RECIPIENT-THING*.

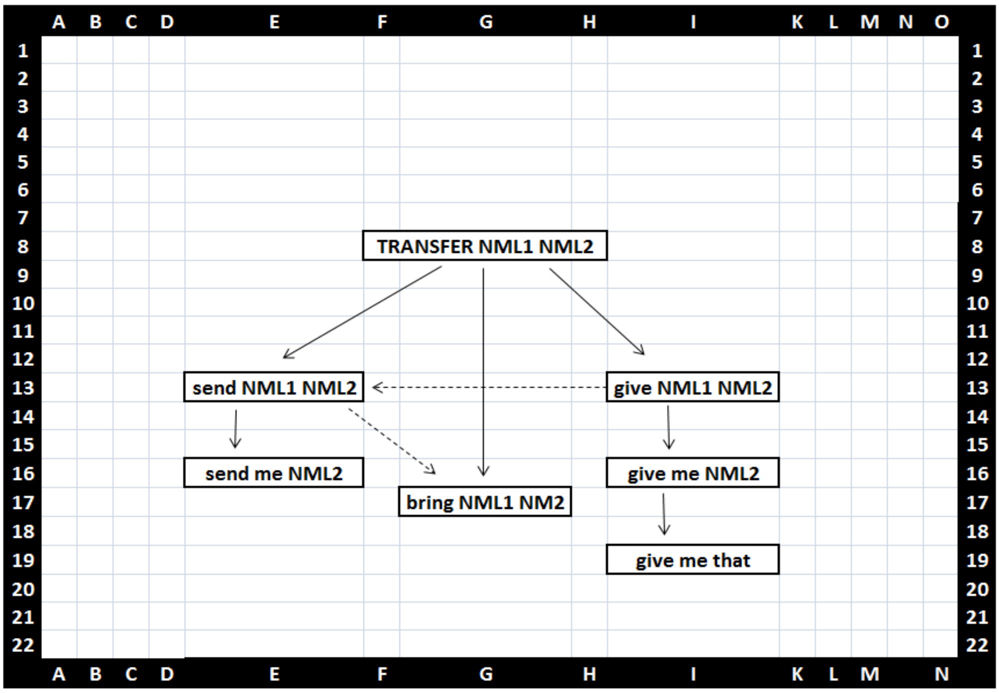


Figure 2.13: the ditransitive complex category 1 of 5: original nucleus. See fig. 2.8 on how to read this and the following figures (fig. 2.14-2.18).

Langacker (2008) appears to suggest that, supposedly, the first usages of verbs of creation in ditransitive frames (33) represented fleeting **extensions** vis-à-vis *TRANSFER-NML1-NML2*, rather than elaborations of it.

33. Rob baked Mary a cake.

The TS *Rob baked Mary a cake* is apprehended as a distorted instantiation of (extension of) the SS *TRANSFER-NML1-NML2*; *a cake* is made available to *Mary* by *Rob* through *baking*. If *bake* starts recurring in a ditransitive frame with enough frequency, *bake-NML1-NML2* may entrench and attain status of unit. Since this use of *bake* is apprehended as an extension vis-à-vis *TRANSFER-NML1-NML2*, even the relationship between the two enters the system (fig. 2.14).

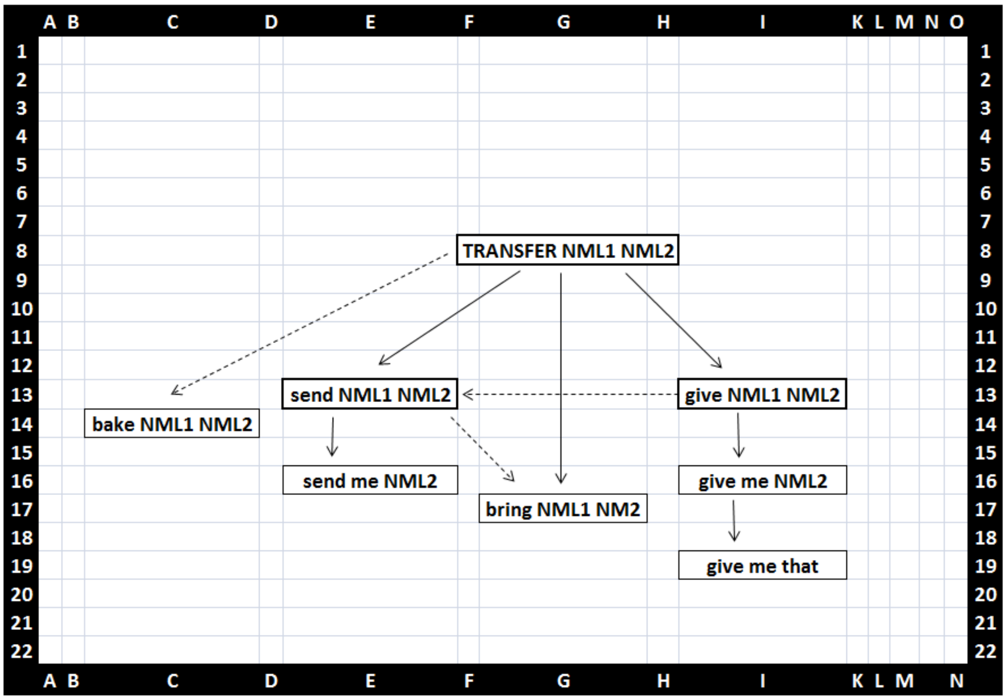


Figure 2.14: the ditransitive complex category; 2 of 5.

The newly attained status of units of *bake-NML1-NML2* as an extension of *TRANSFER-NML1-NML2* may facilitate the use of other verbs in a ditransitive frame (34), as once a unit enters the category, it can serve as the basis for assimilating other expressions to that same category (Langacker, 2008).

34. Mary wrote Mark a long letter.

(34) can be motivated as an extension vis-à-vis either *bake-NML1-NML2* (35) or *TRANSFER-NML1-NML2* (36).

35. [bake-NML1-NML2] ---► Mary wrote Mark a long letter.

36. [TRANSFER-NML1-NML2] ---► Mary wrote Mark a long letter.

If *write-NML1-NML2* became entrenched, it would attain status of unit and enter the system; and so would its relationships with *bake-NML1-NML2* and *TRANSFER-NML1-NML2* (fig 2.15).

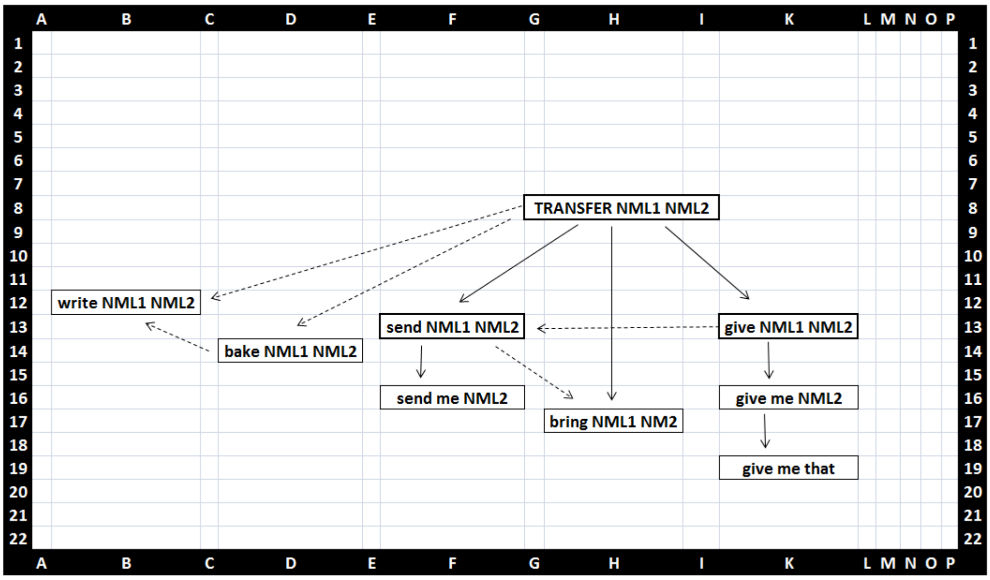


Figure 2.15: the ditransitive complex category; 3 of 5.

To some extent, cases of extension imply that some kind of schematisation is being carried out (Langacker, 2000). If a TS B conflicts with some of the specifications of a SS A, the only way to conceive B as belonging to the same category as A is to filter out differences between the specifications of the two units and conceive some kind of similarity between them. This is possible only through abstracting a unit A', which embodies such a perceived similarity (i.e. A' is a schema that both A and B instantiate) (Langacker, 2000). This is normally represented as in fig. 2.16 (solid and dashed arrows indicate relationships of instantiation and extension, respectively[21]). A' is schematic with respect to A and B, but also represents an extension of A, in that it is obtained by suppressing some of the specifications of the latter.

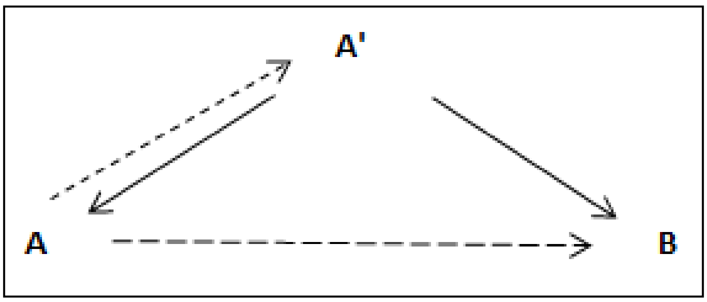


Figure 2.16: schematisation, instantiation and extension.

Hence, [[bake-NML1-NML2]---► [write-NML1-NML2]] is likely to yield the schema *CREATION-NML1-NML2* (fig. 2.17; 8C), which is instantiated by both

[21] Figure 2.16 is identical to Langacker's (2000, 13) fig. 3.

bake-NML1-NML2 and *write-NML1-NML2*. To the extent that both lexically-bound schemas are apprehended as extensions of *TRANSFER-NML1-NML2*, also *CREATION-NML1-NML2* inherits such a relationship (fig. 2.17; see the dashed arrow that links *TRANSFER-NML1-NML2* and *CREATION-NML1-NML2*).

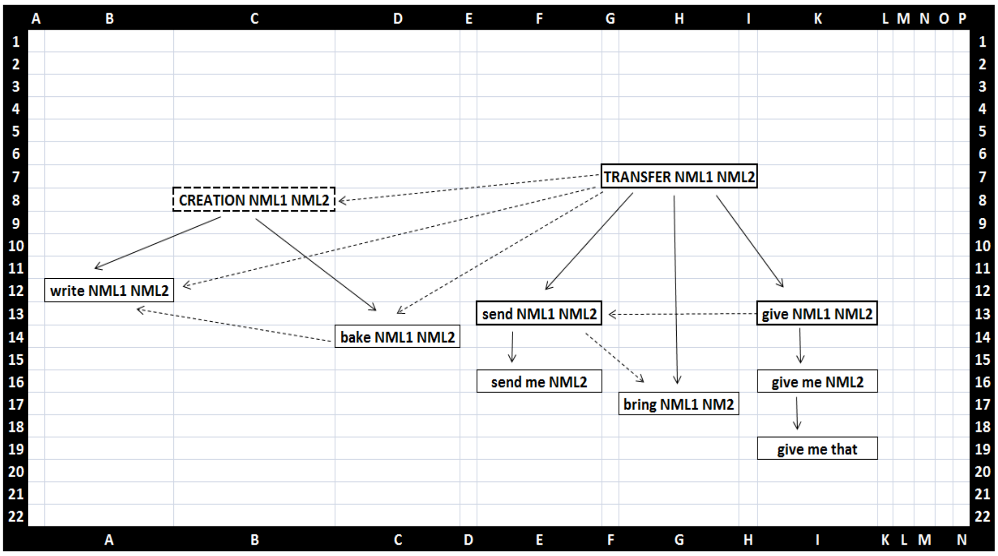


Figure 2.17: the ditransitive complex category; 4 of 5.

Note that *CREATION-NML1-NML2* is enclosed in a dashed box (fig. 2.17; 8C); this indicates that this schema is not entrenched enough to be considered a conventionalised unit retrievable to sanction new expressions. The extent to which a schematic unit is retrievable independent of its instantiations, and hence usable to sanction new expressions, is a function of its entrenchment (and salience). If *CREATION-NML1-NML2*, which is (extemporarily) abstracted when *write-NML1-NML2* is evoked as an extension vis-à-vis *bake-NML1-NML2*, starts gaining cognitive salience, it may become entrenched and attain status of unit (fig. 2.18). The schema can then sanction new usages of the ditransitive frame (37).

37. CREATION-NML1-NML2 $\longrightarrow$ Rob built Mary a house.

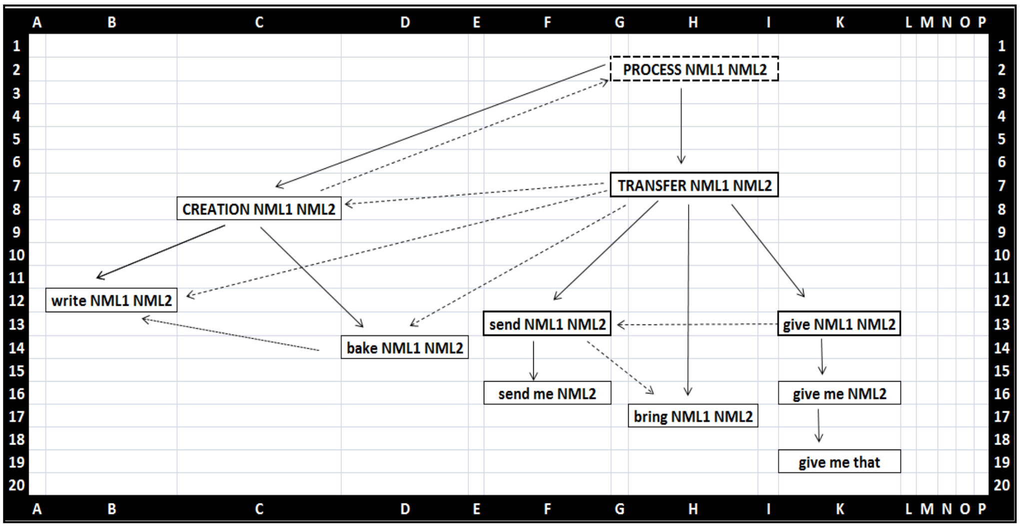


Figure 2.18: the ditransitive complex category; 5 of 5.

TRANSFER-NML1-NML2 and *CREATION-NML1-NML2* might trigger a further schematisation, namely *PROCESS-NML1-NML2* (fig. 2.18; G2-I2), which is schematic vis-à-vis both constructional schemas. The extent to which the new schema can attain status of unit and be evoked independent of its instantiations, will depend on its cognitive salience and degree of entrenchment. Goldberg (1995, 2006) and Langacker (2008) note that such a schema is not available for the sanction (and production) of new expressions and that the ditransitive complex category is not dominated by a single higher-level schema. Indeed, many complex categories (i.e. networks of semantically and formally related constructions; refer back to 2.3.21) do not present a higher-level schema subsuming all the units of the network. Fig. 2.19 reproduces Langacker's (2008, p. 37, fig. 2.2) representation of the lexical item *ring*. The category has two high-level schematic meanings (CIRCULAR ENTITY and ARENA), but there is no higher-level schema instantiated by all units. As already noted, lower-level schemas are likely to be more viable to linguist competence, because higher-level schemas might either not exist at all (as with *ring*) or not be cognitively salient nor entrenched enough to sanction new expressions (as with the ditransitive).

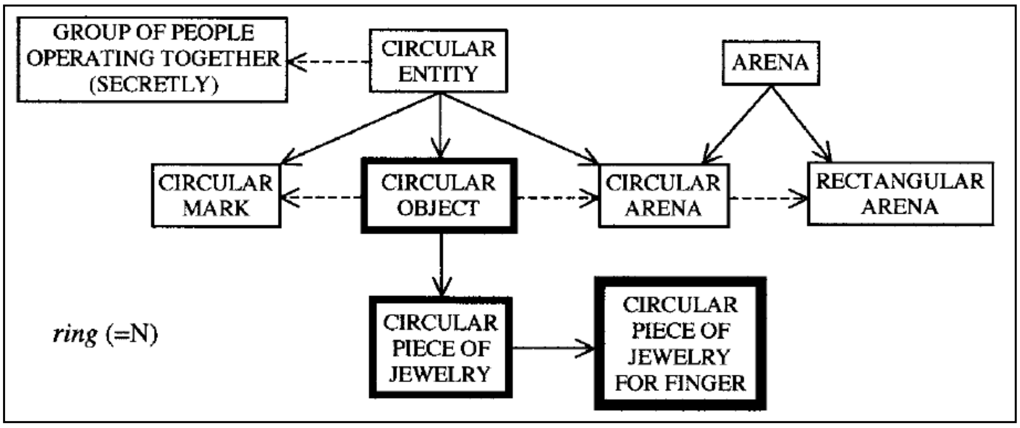


Figure 2.19: the polysemous lexical item *ring* represented as a complex category (from Langacker, 2008, p. 37, fig. 2.2). Dashed arrows indicate relationships of extension. Solid arrows indicate relationships of elaboration/instantiation. Thickness of boxes indicates the degree of entrenchment of the units enclosed in them.

2.6. A UNIFIED ACCOUNT

Unlike traditional accounts of language, which draw a clear line between lexicon (a set of fixed expressions) and grammar (fully general and productive rules), CG assumes the so-called **lexicon-syntax continuum** (Langacker, 2000, 2008).

Language is an inventory of symbolic units, all of which embody generalisations of some kind inferred from any regularity found in **language use**. These units differ in size and specificity and the generalisations they embody are of any kind and scope. Relatively narrow generalisations are needed to yield a lexical item like *cat*, whereas the scope of the generalisations yielding constructional schemas (e.g. *AGENT-PROCESS-PATIENT*) is much broader.

Hence, no clear-cut distinction between what is part of the lexicon (the fixed expressions of a language) and what is not part of the lexicon is hypothesised. Lexical items (*cat*, *dog*), irregular patterns (*go*, *went*, *gone*), formulaic phrases (*it's raining cats and dogs*) and expressions that are fully regular (*sorted*, *give me that*, *I'll do it*) can become entrenched units available as single *gestalts* and posit themselves along a continuum of schematicity and generalisability.

Since the "status of units" of an expression/pattern is a by-product of its entrenchment and since grammatical regularities are constructional schemas that embody those regularities, the system continually shapes and is shaped by language use. That is why these approaches are said to be **usage-based**.

Page intentionally left blank

3.

A USAGE-BASED APPROACH TO LANGUAGE ACQUISITION

Usage-based researchers claim that linguistic knowledge can be described as mastery of a highly interconnected network of symbolic units of any size and degree of specificity. Language acquisition (LA) is about acquiring an inventory of symbolic units and the relationships between them. Children go about such a task by relying on some species-specific skills, such as intention reading, schematisation, pattern finding and cultural learning (Lieven & Tomasello, 2008; Tomasello, 2003, 2006a, 2006b).

Throughout the following chapter[22], the reader is introduced to three main claims about LA and early childhood linguistic representation that characterise the theoretical approach taken here.

a) **Children's language acquisition is piecemeal**

Children start out by acquiring form-meaning pairings (i.e. symbolic units) of any size (*dogs*, *cuddle*, *I eat an apple, give me that*). As they acquire those units, children start analogising across strings perceived as similar on the basis of functional, distributional and semantic similarities, as well as on the basis of the **concrete and recurrent lexical-material** those strings share. Such analogies allow children to develop lexically-specific schemas (*EATER-**eat**-THING_EATEN*) out of the concrete strings they encounter (*I **eat** an apple*; *we **eat** pasta*). The same process of analogy is then further carried on as children start analogising across lexically-specific schemas (*EATER-eat-THING_EATEN*, *I'm-PROCESSing-it*, *HITTER-hit-HITTEE*) on the basis of which adultlike fully schematic units (*AGENT-PROCESS-PATIENT*) are acquired.

[22] Section 3.1 overwhelmingly draws on Dąbrowska (2004), whereas the rest of the chapter summarises and draws on Tomasello (2003, 2006b) and Ambridge and Lieven (2011).

b) Children's language is concrete (i.e. lexically-specific)

Children's initial linguistic knowledge can be described as mastery of an inventory of **independent**, lexically-specific units, which are either **fully lexically-specific** (*give-me-that, dogs*) or **only partially schematic** (*EATER-eat-THING_EATEN*, *that's-a-THING*). Children's and adults' inventories are considered to be very different in nature. Adults' inventories are highly interconnected and their units can have any degree of specificity, from fully-concrete (*give-me-that*) to fully-schematic (*TRANSFER-RECIPIENT-THING*). Children's inventories have fewer units, which are both poorly interconnected and bound to concrete lexical and/or morphological material. Put simply, **children's inventories lack fully-schematic units**.

c) LA is input-dependent

The types of schemas children develop and the generalisations they draw crucially depend on the specific language they encounter. That is, children learn the language they hear (Tomasello, 2003).

3.1. ACQUIRING SYMBOLIC UNITS

In order to acquire symbolic units, children have to map sequences of phonemes onto chunks of semantics in the strings they hear. In order to do so, they have to identify phonological units (i.e. they have to single out words from the speech stream) and map them onto meanings. Jusczyk (1997) reports on various studies which showed that seven-month-olds are capable of singling words out of concrete strings, even when they could not possibly know what those words mean (according to parents' reports). Infants exposed to multi-word sentences containing specific words (*hamlet*, *kingdom*) during training, listen longer to one-word sentences containing those same words than one-word sentences that do not during the test phase. Findings hold in the opposite direction as well; when infants are exposed to single-word sentences (*hamlet*, *kingdom*) during training, they listen longer to multi-word sentences containing those same words than multi-word sentences that do not during the test phase. Thus, children are capable of both learning new words (meant as phonological units) from the speech stream and recognising words they know in the speech stream several months before the onset of meaningful speech (10-12 months).

Once the sound stream has been broken down into units, children have to assign a meaning to each of those phonological units. Initially, children learn chunks of language of any size, from single words to whole sentences, by pairing strings of phonemes to specific meanings. Thus, *dog* is linked to the meaning of a dog and *we eat soup* is linked to a holistic semantic representation of the speaker and others eating soup.

3.1.1. WE EAT SOUP: AN IDEALISED EXAMPLE OF ACQUISITION

Let us suppose that Claire hears *we eat soup* while dining with mum and dad. She is able to break the string into /wi/, /i:t/ and /su:p/ (Jusczyk, 1997) and also understands the communicative intention of the speaker and the social routine of eating (Tomasello, 2003). Claire might then start using *we eat soup* any time she eats soup with her parents. Meanwhile, she might be able to infer the specific meanings of *we* and *soup* from other strings she encounters (*shall* ***we*** *go? The* ***soup*** *is hot*). Similarly, she might experience *we eat soup* in other contexts (eating outside with her grandpa). The interaction of those experiences would allow her to:

a) narrow down the meaning of *eat* to the action of ingesting food, irrespective of, for example, the external setting in which eating occurs and the people with whom she eats.
b) link each phonological chunk (/wi/, /i:t/ and /su:p/) to a meaning.

Once Claire has made steps (a) and (b), she has acquired a mini-grammar of *we eat soup* (i.e. she has parsed it).

Claire might also experience *I eat pasta*. Assuming that she has learnt what both *I* and *pasta* mean, she now has a mini grammar of *I eat pasta*. Other similar sentences might be encountered, learnt and parsed (*daddy eats chocolate*, *you eat vegetables*).

Claire might then start noticing that all these strings are very similar: they all share the lexical item ***eat***, they all map onto the live event of eating and, most crucially, the eater appears before *eat* and the thing eaten appears after it. Hence, she could start drawing analogies across those strings and infer the **lexically-specific schema** *EATER-eat-THING_EATEN*.

3.1.2. LEXICALLY-SPECIFIC SCHEMAS

Units such as *EATER-eat-THING_EATEN* are called **lexically-specific** or **lexically-bound** or **lexically-based schemas** because they contain material (lexical items) that is phonologically specified (*eat*). They are also partially schematic (or partially abstract) in that they contain some elements (called *SLOTS*) which are not phonologically specified (*EATER* and *THING_EATEN*) and which represent form-function abstractions (Lieven & Brandt, 2011). *EATER* has a form, which consists of taking a preverbal position. Its function in the schema is delineating the entity that performs the action of eating. It is an abstraction, as it is the result of the child retaining common features shared by the specific items encountered in preverbal position with *eat* (the fact of being eaters and THINGS in Langacker's (2008) terms), while getting rid of non-shared features (the specific forms which concretely appeared before *eat*). However, insofar as those units are lexically-bound, so are the generalisations they contain: the preverbal argument is an *eater*, not an agent or a subject. Similarly, *eat* is not represented as a verb. Rather, it is a concrete sound-meaning (or signifier-signified) pairing around which the schema is built. *EATER-eat-THING_EATEN* represents lexically-specific knowledge about the verb *eat* and is a semi-formulaic unit that maps onto the life-event of eating. The child learnt that *EATER* and *THING_EATEN* are expressed in pre and post-verbal position, respectively and that she can rely on that schema in order to describe/comment on events of eating.

Two further aspects of lexically-specific schemas need pointing out:

a) Lexically-specific schemas are input dependent

Children develop these schemas from the concrete language they experience. If a child experiences *eat* only in intransitive frames, his/her knowledge of *eat* will be intransitive (*EATER-eat*).

b) Lexically-specific schemas represent the foundations towards productivity

Once a child has acquired the schema *EATER-eat-THING_EATEN*, s/he can evoke it to say things s/he has never heard nor said before (*they eat pizza*).

3.2. LEXICALLY-SPECIFIC (CONCRETE) COMPETENCE

Children's early knowledge of their language is said to be concrete (or lexically-specific) in that it can be described as an **inventory of independent lexically-bound units** (*give-me-THING*, *EATER-eat-THING*). Each of those units is bound to (or built around) some kind of concrete lexical material and has a (unique) meaning and communicative function.

3.2.1. AN INVENTORY OF LEXICALLY-SPECIFIC UNITS

Since children's units are form-meaning pairings which map onto quite specific life events, serve specific communicative functions and are built around specific lexical items, they are said to be **independent from each other**. *KICKER-kick-KICKEE* is not the by-product of an abstract pattern (*AGENT-PROCESS-PATIENT*), rather it is a schema that maps onto the life event of kicking and represents learnt knowledge of *kick* and its frame. Thus, *KICKER-kick-KICKEE* and *EATER-eat-THING_EATEN* are two independent constructions that map onto different life events. Each unit is then a sort of mini-grammar, an island of structural organisation

> "in an otherwise unorganised language system"
>
> (Tomasello, 2003, p. 117).

Fig. 3.1 (overleaf) graphically represents what a child's inventory may look like.

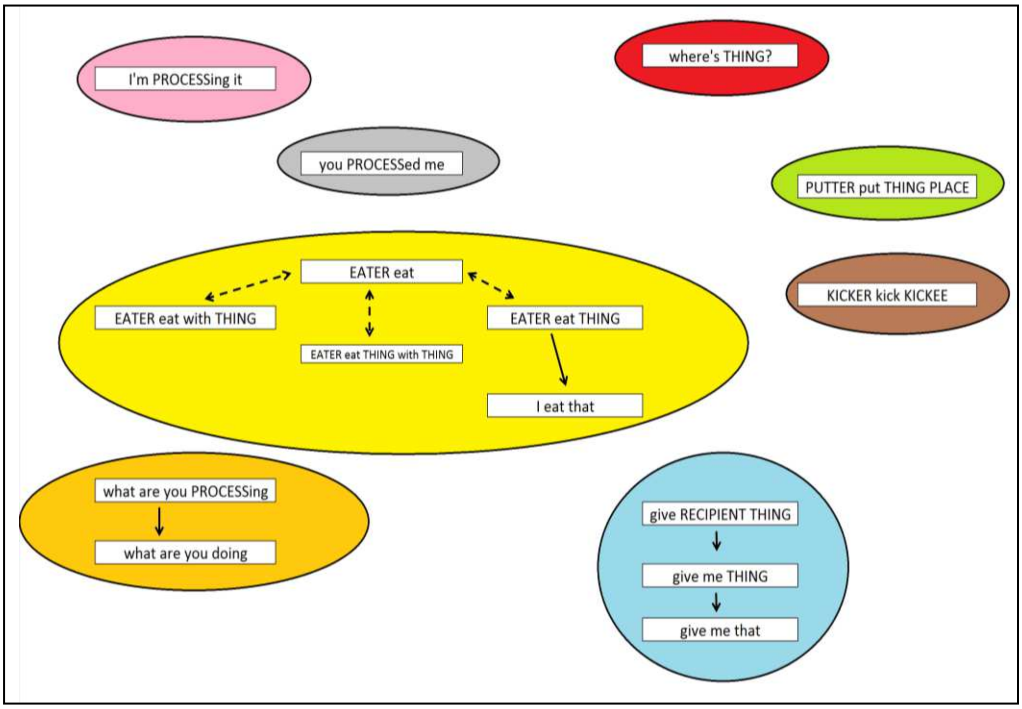


Figure 3.1: Children's inventory of constructions. Each construction can be thought of as a mini-grammar representing lexically-specific knowledge. Symbolic units are enclosed in rectangles in which *small letters* indicate elements that are concrete (phonologically specified) and *CAPITAL LETTERS* indicate slots. Dashed arrows indicate relationships of extension, whereas solid arrows indicate relationships of elaboration.

3.2.2. ADULTS' AND CHILDREN'S NETWORKS

As observed in the previous chapter, both Adults and children have and rely on lexically-specific units. Nevertheless, adults' and children's systems differ in two important respects:

a) Since constructions (units, schemas) are learnt from the input and since children have less linguistic experience than adults, the former's inventories are poorer, with fewer and less entrenched constructions.

b) More importantly, adults' and children's inventories differ in the degree of schematicity of their generalisations (slots), the richness of connections between units and, consequently, the extent to which fully-schematic (abstract) units (e.g. *AGENT-PROCESS-PATIENT*) are entrenched and available to sanction new expressions.

Children develop their lexically-bound schemas by abstracting them from the concrete strings they learn. Those schemas contain slots that represent generalisations that are functional and semantic, rather than syntactic in nature.

Such a functional nature is the key to understanding the difference between adults' and children's linguistic representation.

Slots in children's schema

> "are defined by the role they play in the schema"
>
> (Tomasello, 2003, p. 124)

and hence they exist solely in relation to the schema of which they are part. *THING_EATEN* in *EATER-eat-THING_EATEN* is nothing but something that can be eaten and KICKEE is just something that can be kicked.

Adults' generalisations, on the other hand, are syntactic in nature, as they belong to units composing much wider and more richly interconnected networks than those of children.

As previously pointed out in chapter 2, information in a speaker's network may have a high degree of redundancy, in that the same pattern can be represented at different levels of specificity (*SVO*, *AGENT-PROCESS-PATIENT*, *EATER-eat-THING_EATEN*, *EATER-eat-that*, *I-eat-that*). Crucially however, information at the most schematic level is putatively available to sanction new expressions. Hence, adults may well store lexically-specific units such as *KICKER-kick-KICKEE* and *EATER-eat-THING_EATEN*, but these are linked to one another by relationships of categorisation and composition and, more crucially, they are linked to more abstract patterns. Consequently, both *KICKEE* and *KICKER-kick-KICKEE* will hold relationships of elaboration with more abstract units (*PATIENT* and *AGENT-PROCESS-PATIENT*, respectively:
PATIENT → KICKEE and *KICKER kick KICKEE* → *AGENT-PROCESS-PATIENT*).

Children's inventories lack these superordinate fully-schematic layers, in that each unit/construction is tied to some specific lexical material and represents lexically-specific knowledge that is independent from other islands (constructions/units). Fig. 3.2 graphically represents the different types of representations that may underline adults' (fig 3.2a) and children's (3.2b) knowledge of the ditransitive construction[23].

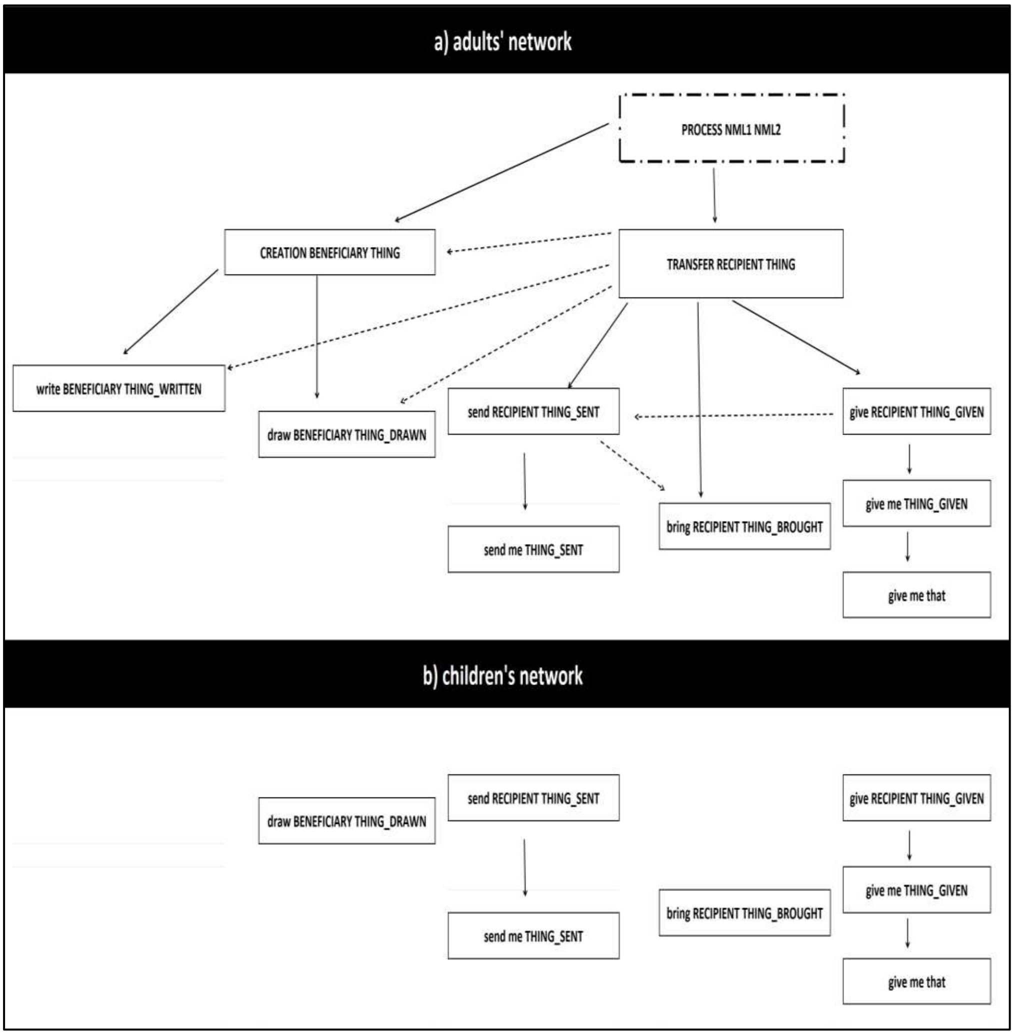


Figure 3.2: Adults' (a) and children's (b) representation of the ditransitive construction. Children lack the more schematic layers and their units are fewer and much more poorly interconnected. Dashed arrows indicate relationships of extension, whereas solid arrows indicate relationships of elaboration. Symbolic units are enclosed in rectangles. *Small letters* indicate lexically-specific elements, whereas *CAPITAL LETTERS* indicate slots. Thickness of rectangles indicates the degree of entrenchment of the units contained in them.

3.2.3. IMPLICATIONS FOR ACQUISITION

The hypothesis that children's knowledge of their language is tied to lexically-specific patterns acquired from the input and that initially such patterns are

[23] As Langacker (2008) points out, the network is a metaphor and must not be taken too literally. Nevertheless, it serves the expositive nature of this chapter well and hence I shall continue to adopt it throughout this thesis (see Goldberg (1995) and Langacker (2008) for discussions on the limits of the network metaphor).

independent form each other, leads to the following predictions (or consequences):

a) **Uneven development and complexity**

If the child's knowledge of *hit* is independent from knowledge of, for example, *draw*, what s/he learns about the former will be unrelated to what s/he learns about the latter; since children's constructions are independent from each other, children will develop them independently. Thus, the child might learn different verbs in different frames (*X-draw*, but *X-hit-Y*) and also expand those lexically-specific units differently, resulting in an uneven complexity of frames for different verbs (*X-draw-Y*, but *X-hit-Y*, *X-hit-Y-on-Z*, *X-hit-Y-with-K*).

b) Input-dependent development

Children draw their inventories from the concrete language they experience. Hence, children's language should mirror the specific input to which they are exposed and the way their constructions develop should be input-dependent. Consequently, the different developments of *draw* and *hit* above are likely to be by-products of the child having experienced *hit* in a much wider set of constructional frames. Because children tend to "stick to the input", it is often stated that they are **conservative** learners.

c) Piecemeal development

Since the linguistic generalisations children draw are both lexically-specific and input dependent, language development will be a slow and gradual process. Constructions/units will grow bit by bit on an item-by-item basis, rather than display across-the-board improvement. Thus, what a child learns about *kick* is not (fully) generalised to other verbs.

3.3. CONSTRUCTING A LANGUAGE

Developing adultlike linguistic competence is about acquiring an inventory of highly interconnected symbolic units varying in size, specificity and generalisability. What are traditionally regarded as grammatical patterns/rules are actually sets of complex fully-schematic constructions (the ditransitive construction, the transitive construction, the wh-question construction, etc.) that can be thought of as meaningful templates or, in Langacker's (2008) terms, constructional schemas. The path towards the acquisition of such units starts out with learning chunks of language of any size, from single words (*dogs*, *cat*) to whole sentences (*give me that*, *where's mum?*) that are initially **fully concrete** in that they are symbolic units whose phonological pole is fully specified. At early stages, more complex units are learnt as big words which link a string of phonemes to a (holistic) meaning. Once children have "parsed" or analysed (at least parts of) those strings (*give me that*) into their components (*give_me* < *that*), and mapped at least some of such components onto real life entities or communicative functions, they can start drawing analogies across similar strings (*give me that*, *give me my hat*) and develop lexically-specific schemas (*give-me-THING*). As children become better parsers and acquire more and more units, they can start drawing analogies across a wider range of strings (*give mum her keys*, *I give you some chocolate*) and develop their schemas in abstraction/schematicity (*give-me-THING* develops into *give-RECIPIENT-THING*), gradually relying on (abstracting) constructions that are more and more schematic (*give-RECIPIENT-THING* develops into *TRANSFER-RECIPIENT-THING*).

3.3.1. DEVELOPING SCHEMATIC UNITS

Dąbrowska (2000) followed the development of several types of syntactic questions in an English-speaking girl (Noemi) from 1;6 to 3;8. The child started off by imitating *what's mommy doing?*, which she repeated several times in the following weeks. Noemi had substantially learnt a formula, that is, a frozen string that had word-like status (a fixed sequence of phonemes linked to a meaning). Around the age of 2;0.18 she uttered *what's donkey doing?*, *what's toy doing?* and *what's Nomi doing?* This suggests that the formula *what's mommy doing?* had developed into the schema *what's-THING-doing?* That is, the element *mommy* had developed into a schematic slot (*THING*). By relying on this schema, Noemi could therefore fill the slot *THING* with various lexical items in order to produce new utterances. Her language had become more productive and more flexible. At age 2;1.0 she uttered *what's George saying?* and *what's mommy holding?* This suggests that *what's-THING-doing?* had undergone further development in abstraction. Noemi was then possibly relying on the schema *what's-THING-PROCESSing?* whose newly formed slot

PROCESS allowed her to enquire into the kind of activity in which someone was engaging.

The overall picture is consistent with the hypothesis that children start out by imitating the input, therefore acquiring frozen formulas that **gradually** develop into more and more schematic units.

3.3.2. PIECEMEAL AND CONTINUOUS DEVELOPMENT

Before children start developing lexically-specific schemas, around 16-18 months of age, they start combining frozen strings (holophrases) and words. For instance, they might say *apple chair* or *chair apple* to draw adults' attention to the fact that an apple is lying on a chair. Those sentences are evidence that children can partition scenes into different entities and map them onto linguistic symbols (Tomasello, 2003).

At around 18 months, children start showing some consistent patterns in which a recurring word is combined with a limited set of semantically related words. Such patterns, called **pivot schemas** (Braine, 1976) or positional patterns or low-scope formulas (Tomasello, 2003; Ambridge & Lieven, 2011), show a consistent and a fixed linear order (*Kick-X*, *Bye-bye-X*, *X-gone*, *more-X*). Initially, pivot schemas co-exist with less mature (*chair apple*) strategies of word combination.

For instance, during her two-word stage (16-18 months) Tomasello's (1992) daughter would consistently use certain verbs (*catch*, *see*) with VO sequences (*see-THING_SEEN*, *catch-THING_CAUGHT*). Yet, during the same developmental period some other verbs were as likely to appear with adultlike orders (*get THING_GOTTEN*) as to appear in ungrammatical sequences (*THING_GOTTEN get*). These findings indicate that the child had learnt pivot schemas for some items (*catch-THING_CAUGHT*) that allowed her to utter adultlike sentences. However, since she had not developed more schematic units (*PROCESS*, *AGENT*, *PATIENT*) to which she could assimilate all of her verbs and since each pivot-schema represented lexically specific knowledge (of, for example, *see*), she could not generalise what she learnt about *see* (the VO order) to other verbs (e.g. *get*). As a result, her usage of verbs for which she had yet to learn specific patterns (e.g. *get*) was inconsistent.

Later on in development, pivot-schemas gradually develop in complexity (length) and schematicity. When Tomasello's (1992) daughter started producing longer sentences, she did so by expanding her previous pivot schemas and formulaic strings by means of small steps. For instance, she started adding new arguments to attested pivots (e.g. adding a *HITTER* to *hit-HITTEE* yielded *daddy hit tennis*, presumably developing the schema *HITTER-hit-HITTEE*). She also started coordinating existing pivot-schemas (e.g. *X-stuck* and *stuck-on-Y* yielded *X-stuck-on-Y*; *marshmallows stuck on there*; Tomasello, 1992, pp. 230-243).

Importantly, during the same developmental period (18-24 months) there was great discrepancy in how (and whether) she used and marked various arguments with different verbs. For instance, the child used *cut* in only one (subjectless) frame (*cut-X*), whereas *draw* was attested in eight different frames (*X-draw-Y*, *X-draw-Y-on-Z* and so forth; Tomasello, 2003, 1992). Locatives were marked by prepositions with some verbs (*stuck on bowl*), but not others (*spill it couch* meaning *spill it on the couch*; Tomasello, 1992). At any given developmental point, the best predictor of how a verb would be used (in terms of the frames in which it appeared and how arguments were marked) was **not** how the child used other verbs during that same developmental point, which would suggest an across-the-board, verb-general competence. Instead, the best predictor of how and whether she would mark the arguments of a specific verb (i.e. the frames in which she would use it) was how she previously used that same verb, which suggests continuity on a verb-by-verb basis, but discontinuity and independence across verbs.

Such findings were confirmed by McClure, Pine and Lieven (2006), who investigated the language produced by ten English-speaking children from 1;10-2;0 to 2;3-3;0 during MLU phase 1 (M=1.46) and phase 2 (M=2.43). They compared the complexity of frames (number of arguments) of two kinds of verbs during MLU phase 2: those that had been used during phase 1 (old verbs) and those that appeared for the first time during phase 2 (new verbs). The rationale is that, if children's knowledge of argument frames is verb-general (fully-schematic), they should be able to use different verbs in similar frames. Conversely, if their knowledge is lexically-specific, verbs with which children have more experience (old verbs) should appear in more complex frames. Since children are likely to have encountered those *old verbs* more frequently than *new verbs,* they should have had more opportunities to learn a wider variety of lexically-specific patterns built around the former than around the latter. Old verbs were indeed combined with significantly more arguments (M = 1.14) than new verbs (M= 1.0) ($t(9)= 2.26$, $p<.05$, one-tailed; McClure et al., 2006, p. 709).

These findings suggest that development is piecemeal, incremental and continuous, but it is so on an item-by-item basis. Each lexically-specific schema maps onto its own meaning, with little or no connection to other schemas, and develops independently of other constructions.

3.3.3. ON CONSTRUCTIONAL ISLANDS

The child's grammar in Tomasello's (1992) study appeared to be organised around verb-specific constructions, which led Tomasello to hypothesise that children's early language was structured as

> "an inventory of independent verb-island constructions that pair a scene of experience and an item-based construction, with no structural relationships among these constructional islands"
>
> (Tomasello, 2003, p. 121).

This hypothesis is known as **Verb Island Hypothesis** (VIH, henceforth).

However, McClure et al. (2006) also found that the first occurrences of specific verbs during MLU phase 2 presented more complex frames than the first occurrences of (other) specific verbs during MLU phase 1 ($t(9)=2.72$, $p<.05$, one-tailed; McClure et al., 2006, p. 711). McClure et al. (2006) observe that such findings are not consistent with the VIH, as they suggest that children could somehow transfer some of their knowledge of older verbs to newly acquired ones. Following findings of previous studies (e.g. Pine & Lieven, 1993; Lieven, Pine & Baldwin, 1997; Akhtar, 1999; Childers & Tomasello, 2001), McClure and colleagues hypothesised that one of the possible explanations for their outcomes was that children also build/infer lexically-specific schemas that are verb-general (e.g. *I'm-PROCESSing-it*), whose slots represent generalisations across different specific verbs (as in the case of Noemi's *what's-THING-**PROCESS**ing?* in Dąbrowska, 2000). Indeed, a more qualitative analysis of children's sentences showed that all children could be said to have developed the schemas *I-PROCESS* and 90% of them also appeared to rely on the schema *PROCESS-it*.

> "Slots develop as a result of variation in a repeated string […]. If the child can insert a novel item into the slot, this is evidence that a form to function abstraction has been made and schematisation has occurred"
>
> (Lieven & Brandt, 2011, p. 285).

If a pattern is recurrent enough, it entrenches and achieves "status of unit", becoming retrievable to sanction new expressions. The recurring of patterns such as *I eat*, *I sleep, I go* might help the formation of the schema *I-PROCESS*, whose slot can be paraphrased as "action I do". Once a child has learnt such a formula, s/he can speak about newly learnt actions (*spin*) in a way that may have not been experienced before (*I spin*).

Thus, children do not build their constructions around verbs only. Rather, they acquire the schemas that are frequently instantiated in their ambient language and that meet their communicative needs (Ambridge & Lieven, 2011, pp. 201-202).

The take-home message is that early language is an inventory of independent lexically-specific constructions abstracted from the concrete input that children experience. Those constructions may be built around any kind of recurring and specific lexical material shared by the concrete strings that instantiate them (*I'm-PROCESSing-it*, *EATER-eat-THING_EATEN*). The degree of the generalisations of those units is however lexically-bound. In such a limited degree of abstraction lies the main difference between adults' and children's linguistic knowledge.

3.4. EVIDENCE FROM ITALIAN

D'Odorico, Fasolo, Cassibba and Costantini (2011) investigated thirty minutes of spontaneous production of forty-five Italian-speaking children, aged 1;11-3;0, divided into two MLU groups (group 1, MLU=1.50-1.97; Group 2, MLU=2.03-2.98). Of the 160 verb types in the 2223 children's utterances analysed, ten verbs accounted for 50% of all tokens. This is consistent with the idea that children's inventories are much smaller than those of adults and that, initially, children's language is based around a limited set of well-rehearsed lexically-specific patterns. Children also produced the majority of verbs in one, and only one frame, whether they belonged to group 1 (67%) or group 2 (57%). Only 2% (group 1) and 5% (Group 2) of verbs were attested in three or more different frames. Hence, children did not appear to be capable of using different frames with different verbs, which suggests that frames were learnt on a verb-by-verb basis. The implication is that the frames in which verbs appeared were constructional islands of limited scope. Mixed verbs, which could be used both transitively and intransitively, were overwhelmingly used in a single frame by Group 1 children (2% intransitively and 68% transitively = 70%), whereas 60% of mixed verbs were used both transitively and intransitively by Group 2 children. Again, this is consistent with the hypothesis that, as children gain more linguistic experience, they expand and develop their lexically-specific patterns. More advanced children had more opportunities to learn that those mixed verbs could be used both transitively and intransitively and so Group 2 children used them in both ways nearly twice as much as Group 1 children.

Pizzuto and Caselli (1994) analysed the development of verbal morphology in three Italian-speaking children (from 1;4 to 3;0). In an initial phase (up to 1;11) all children combined specific verb roots/stems with only one person-number marker. After that phase, roots/stems slowly started to be combined with two or more inflections. By the end of the study, for each child's corpora about 50% of stems (44%-53%) had been combined with only one specific person-number marker, whereas only 27-33% of stems appeared with three or more. The scholars also found that different person-number markers (e.g. PRS.1.SG) were learnt at different ages and in different orders by different children. The implication is that children do not have an across-the-board mastery of verb paradigms. Rather, they initially combine some roots/stems with some specific inflections and some other roots/stems with other specific inflections. Thus, even morphological development is a piecemeal process that starts out on an item-by-item basis. Only gradually do children learn to generalise their root-specific knowledge of person-number markers to several different roots.

These studies indicate that initial linguistic competence is not an across-the-board knowledge and, instead, it is describable as a set of lexically-bound specific constructions, be they at the word (morphological inflections and verb roots) or clause (syntactic frames) level.

3.5. THE TRACEBACK METHOD

UB scholars claim that speakers "build" their sentences by juxtaposing and superimposing units of their inventory. Children rely on lexically-bound units that are learnt from the concrete language they themselves have previously experienced. It follows that what children say (*mum*, *that's a pig*) should be accountable for in terms of fully (*mum*, *pig*) and partially (*that's-a-THING*) lexically-specified units that:

a) are superimposed and/or juxtaposed together
b) could have been learnt from the concrete language that children themselves have previously encountered (*give* ***mum*** *a cuddle*, ***that's a*** *car*, ***that's a*** *sheep*, *there's* ***a pig*****)**

In order to investigate such a hypothesis two "types" of data are needed:

a) A sample of children's language during a given point in development (*that's a pig*) to be analysed in terms of lexically-specific units (*that's-a-THING*).
b) A sample of the same children's own previous linguistic experience (*that's a car*, *that's a sheep*) from which those units (*that's-a-THING*) could have been learnt.

Lieven and colleagues (Lieven, Behrens, Speares & Tomasello, 2003; Lieven, Salomo & Tomasello, 2009; Dąbrowska & Lieven, 2005) developed the **traceback method** that specifically investigates such a research question. Because the original research to be presented (part II) analyses the spontaneous production of an Italian-speaking child adopting such a method, this chapter walks the reader through some of its key features and results.

3.5.1. SAMPLING AND CORPORA

Children's spontaneous production is sampled for about five hours a week for six weeks in a row and the dataset collected is divided into two corpora:

a) a **Test Corpus** which consists of the last two hours of recordings (i.e., the last day or last two days of the study)

b) a **Main Corpus** which consists of all remaining recordings (about twenty-eight hours)

The **test corpus** represents a picture of children's language at a given point in development. The analysis is about the language attested in the test corpus and whether it can be accounted for in terms of lexically-specific units that children have encountered during their previous linguistic experience. The **main corpus** represents the children's previous linguistic experience and is the *locus* where the instantiations of the putative units on which the children relied to produce their **test corpus** sentences can be found.

3.5.2. ANALYSIS

Each multi-word sentence type uttered by children in the **test corpus** is identified and traced back to its putative **component units**. Component units are words/strings of words attested in the **main corpus** that share lexical material with a specific test corpus sentence (called **target sentence**). Once a target sentence has been traced back to its putative units, it is derived by superimposing and juxtaposing them.

For instance, let us assume that (1) is the **target sentence** to be analysed and (2)-(4) are the main corpus strings containing its component units. The latter are reported in *italics* and shared lexical material is indicated in both ***bold and italics.***

1. Now I want the big box of sweets.
2. *a. **I want** my book.*

   b. ***I want*** *ice-cream.*
3. I'll get you ***the big box of sweets***.
4. Mum is tired, ***now***.

(2a-b) and (1) share the string *I want*, which is followed by NPs. Those NPs have the function of specifying what is wanted. That is, (2a), (2b) and (1) instantiate the schema *I-want-THING*. Hence, *I-want-THING* is considered one of (1)'s component units. The other component units are the fully phonologically-specified strings *now* (4) and *the big box of sweets* (3). The method derives (1) by applying superimposition and juxtaposition, as illustrated in fig. 3.3.

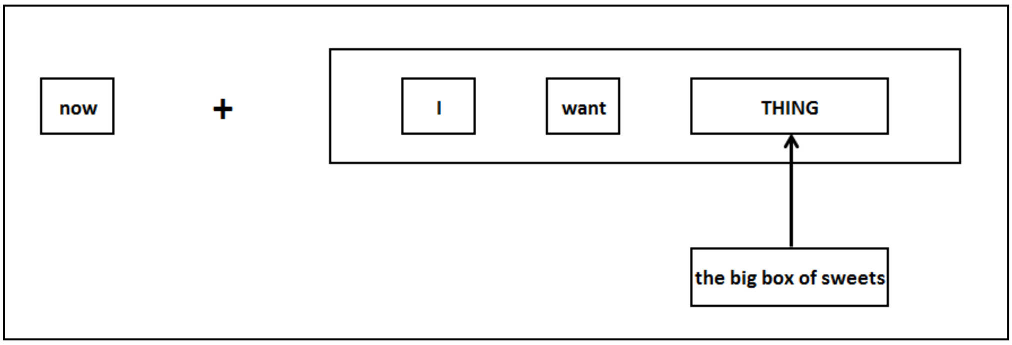


Figure 3.3: the traceback method: deriving *Now I want the big box of sweets* by superimposing (arrow) *I want THING* and *the big box of sweets* and by juxtaposing *now* (+).

The rationale behind such a design lies in the assumption that children's grammar can be described as an **inventory of lexically-specific constructions**. Children draw on their own inventories in order to produce new utterances. Since these schematic constructions are bound to, or built around, specific lexical items, sentences children utter ought to share some kind of concrete lexical material with the schemas they instantiate. Consequently, what a child says (*I want that*) must be traceable back (or linkable) to other previously encountered sentences (*I want this*) which also instantiate the same lexically-specific schema (*I-want-THING*). In order to support such a hypothesis, it must be proven that most of what children say is not "brand new" and is accountable for in terms of lexically-specific units (be they lexically-specific schemas or fixed strings).

3.5.3. RESULTS

Dąbrowska and Lieven (2005) investigated the syntactic questions produced by two English-speaking children when they were two and three years old. They found that 91% of questions in both children's test corpora at age two (e.g. *can he park?*) could be accounted for by superimposing and juxtaposing lexically-specific units (*can-he-PROCESS?*; *can-THING-park?*), with 80% to 91% of tracebacks requiring zero (exact repetition of a string attested in the main corpus) or one operation. Tracebacks at age three were nearly as successful (87%-88%), though more frequently (21%-43%) the analysis required two or more operations to derive a target sentence.

Lieven et al. (2009) investigated the entire production attested in the test corpora of four English-speaking children aged 2;0. For one of them (B.), three further corpora were also analysed (at age 2;2, 2;6 and 2;7). They found that 79% to 93% of sentences could be derived by superimposing and juxtaposing component units attested in children's own main corpora, with 58% to 92% of tracebacks requiring no more than one operation.

In both studies, there was a small proportion of sentences that were classified as **syntactic fails** (0% to 18% across the two studies). These were attested whenever the method failed to identify a lexically-specific schema to which the target sentence could be traced back. The interpretation of those fails is very much a matter of the theoretical framework one chooses to follow. Nativist researchers would probably claim that they constitute evidence that children's underlying representation is abstract. UB scholars may point out that the sampling regime adopted by these studies, as dense as it may be, captures only 5% to 10% of children's linguistic experience (Tomasello & Stahl, 2004; Lieven & Behrens, 2012). It could therefore be argued that it is likely that researchers simply failed to capture on tape the models (schemas) out of which those sentences that were classified as syntactic fails had been built.

Importantly however, despite the differences in the size of the corpora, the types of constructions investigated and some slightly different design choices[24], results are very consistent. The overwhelming majority (79% to 93%) of children's sentences can be accounted for in terms of lexically-specific units that children could have learnt from the concrete language they experienced (i.e. from the strings attested in the main corpora). This implies that children's language can be characterised as concrete, in that it is built on narrow-scope, input-dependent and lexically-bound generalisations.

[24] For instance, in Dąbrowska and Lieven (2005), in order to be classified as an available component unit, a putative precedent of a target sentence had to be attested twice or more in the main corpus, irrespective of who uttered it (children or carers). Lieven et al. (2009) traced sentences back to anything children, and children only, uttered at least once in their own main corpora.

3.6. MORE ON CHILDREN'S LINGUISTIC REPRESENTATION

The claim that children's language is input-dependent and that children are conservative learners is often misinterpreted as a (implicit) claim that children are not productive (or innovative, or creative) and that they cannot go beyond the specific input they hear. Clearly, productivity and ability to go beyond the input are related and often indistinguishable issues; in order to go beyond the input, children must be creative (or productive, or innovative). In the next sub-sections, I aim to clarify that UB scholars neither claim that children never go beyond the input nor suggest that they are not creative/productive. I therefore shall discuss how those two issues are accounted for by a lexically-specific view of early childhood language.

3.6.1. GOING BEYOND THE INPUT

Children build their inventories of (lexically-specific) constructions by learning symbolic units from the input and then they rely on their own inventories in order to produce new utterances; hence they are said to be **conservative**. Stating that children are conservative does not mean that children never attempt to go beyond what they have heard (and know). Were it so, their language would never develop; after all, learning a language implies learning to say things that have never been said before.

The crucial point is that when they go beyond the adult model, their behaviour may (or may not) result in ungrammatical sentences. Tomasello (1992) showed that his daughter had developed lexically-based patterns for some verbs (*catch-THING_CAUGHT*), but not for others (*get*). Crucially, when communicative needs pushed her to go beyond what she knew (the lexically-specific patterns acquired), she still attempted to use those verbs for which specific patterns had yet to be learnt (*get*). Her speech however, was as likely to be grammatical (*get this*) as ungrammatical (*this get*). The extent to which children are linguistically adultlike is the extent to which they have learnt the relevant patterns.

In conclusion, children **do** go beyond the input, but the extent to which they manage to produce adultlike sentences when they do so is probabilistic at best.

3.6.2. LOW-SCOPE AND FULL PRODUCTIVITY

When children learn lexically-specific schemas (*give-me-THING*), they are learning productive formulas that allow them to say new things, that is, SLOTS allow linguistic productivity. Once children start developing such constructions, the best examples of the knowledge underlying their language are represented by cases of **functional coercion** (Tomasello, 2003). Functional coercion is the process through which children fill slots of lexically-bound schemas with fillers that would be inappropriate in the adult language.

Dąbrowska (2004, p. 162) reports on the following example from Clark (1974):

5. I want I eat an apple.

Here the child appears to fill the slot of a putative schema *I-want-X* with the putative formulaic string *I eat an apple*. Slots are developed by drawing generalisations from similar strings. The nature of those slots in children's early language is functional and semantic, rather than syntactic (Tomasello, 2003; Dąbrowska, 2004). As Lieven and Brandt (2011, p. 283) point out, the form-function mapping that is at stake is the mapping of a form onto a "(child identified) function".

In (5), the child probably relies on the putative schema *I-want-WHAT_I_WANT*: what he wants is to eat an apple. He also probably learnt to describe the event of him eating an apple by using the formula *I eat an apple*. Since the scene/meaning described by *I eat an apple* corresponds to what he wants, he superimposes it over the slot of the schema *I-want-WHAT_I_WANT*. The outcome is (5).

All constructions can be represented at different levels of specificity and when speakers express their communicative needs, they can activate any of them. Let us suppose that for the sentence at stake (*I want to eat an apple*) both adult and child activate the schema *I-want-X*. The adult's sentence is grammatical (*I want to eat an apple*), the child's is not (*I want I eat an apple*). Why?

The difference lies in the kind of schema representation the two have. More precisely, it lies in the scope of the slot (i.e. in its degree of abstraction). Whereas the adult is likely to have activated the schema *I-want-VP* (as opposed to *I-want-PATIENT*), the child has activated the schema *I-want-WHAT_I_WANT*. The adult's slot is syntactic, the child's slot is functional. The difference is a matter of degree; adults rely on representations which have a higher degree of abstraction and hence a different kind of productivity. Adult productivity is the by-product of a fully-schematic (syntactic) representation. In the case of children however, productivity is tied to lexically-specific and semantic (low-scope) generalisations. From now on, I shall refer to the former kind of productivity as **full productivity** (FP) whereas the latter kind of productivity will be referred to as **low-scope productivity** (LSP, henceforth), (see also Lieven & Brandt, 2011; Ambridge & Lieven, 2011).

3.7. IS NATURALISTIC DATA ENOUGH?

Thus far, the claim that children's language is rooted in low-scope, lexically-specific schemas has been backed up by evidence collected through naturalistic studies. However, as consistent as it may be, naturalistic evidence could be misleading in two ways: it could either underestimate or overestimate children's grammatical knowledge.

3.7.1. LEXICALLY-SPECIFIC KNOWLEDGE

The problematic status of syntactic fails in traceback studies is a case in point. On the one hand, one could argue that they are evidence that children rely on abstract patterns (possibly even innate principles and constraints), as the data in hand indicate that those sentences cannot be accounted for in terms of lexically-specific patterns. However, such a conclusion might overestimate children's grammatical knowledge. It is worth pointing out that some (up to 5%) of the sentences analysed by traceback studies could not be successfully traced back because they contained a lexical item that was not classified as a unit. However, it does not seem controversial to say that if a child knows a word (e.g. *cat*), it is because s/he has encountered it. What then is to prevent researchers hypothesising that, in the same way as some lexical items have "escaped" sampling (which captures only 5% to 10% of a child's average linguistic experience), so might some of the (lexically-specific) syntactic patterns children used in their test corpora[25]?

[25] See Dąbrowska and Lieven (2005) for similar reasoning.

At the same time however, the fact that children's sentences can be accounted for in terms of lexically-specific patterns does not rule out that those same sentences could have been generated through abstract patterns (see Dąbrowska & Lieven, 2005, p. 464). Young children want to engage in the same activities over and over and like to talk about the same things over and over. They also experience fairly routinised communicative interactions and often have limited MLU and vocabulary (Lieven et al., 2003, 2009; Tomasello, 2006b). Therefore, one may argue that children have a fully-fledged (fully-schematic) grammatical competence, but the interaction of these factors limits the degree of novelty of their sentences and masks such an abstract knowledge under a veneer of lexical specificity.

3.7.2. STATISTICAL PROPERTIES OF LANGUAGE

A subtler reason for which naturalistic corpora could lead researchers to underestimate children's productivity is that adults do not exploit all the distributional possibilities of their language either.

Pine and Lieven (1997) investigated determiner-noun sequences (*a N*, *the N*) in a longitudinal study that followed eleven English-speaking children from 1;0 to 3;0. For each child, they calculated a measure of overlap between the two determiners (the proportion of nouns that were used with both determiners as a function of all nouns that were combined with either one). Seven children used some nouns exclusively with *the* and other nouns exclusively with *a* (0% overlap). Five children had overlap rates ranging from 8% to 44.4%, but none of those figures was significantly different from zero after multiple-test adjustments (Pine & Lieven, 1997, pp. 129-130). Such an uneven distribution suggests that children did not have an abstract category *DETERMINER* and that

> "Children may be rote-learning Determiner + Noun sequences on an instance-by-instance basis".
>
> (Pine & Lieven, 1997, p. 131)

Yang (2009) fiercely warns against assessing children's underlying linguistic knowledge solely on the basis of superficial uneven productivity. Relying on Zipf's law (Zipf, 1949), he notes that the frequency of a lexical item in a given corpus is inversely proportional to its frequency rank in the same corpus. As a result, only a few words will be very frequent and many will occur only once. Hence, those words that appear only once will necessarily be combined with only one determiner. Moreover, particular nouns (*bath*, *toilet*) will overwhelmingly tend to be combined with only one type of determiner (*a bath*, *the toilet*), as a consequence of pragmatic and non-linguistic factors. In his analysis of determiner-noun sequences in the Brown corpus (Kucera & Francis, 1969), which draws on written English, he found that four of the children followed by Pine and Lieven (1997) had overlap values (25%-44%) that were higher than the one attested in the adult written corpus (25.2%). Furthermore, the nouns which appeared with both determiners displayed a strong tendency to be associated with either *a* or *the* (2.86:1). Only 12.5% of nouns displayed equal rates of *a* and *the*. Clearly, if even adults show a limited productivity, one could not expect children to show full productivicty.

The logical conclusion is that one cannot take unevenness of distribution as a measure of productivity, since language patterns themselves quite naturally present skewed distribution.

3.7.3. NOT ENOUGH

Naturalistic data sometimes provide some errors, such as *I want I eat an apple*, which help researchers to tell LSP and FP apart. However, children appear to be, on the whole, cautious speakers and these errors are rare[26]. Consequently, in absence of many (clearly) innovative usages, it is often difficult to unravel the underlying competence children are putting into the task of LA. Naturalistic corpora can tell us only part of what children say and hear, but do not provide evidence of what they do not or cannot produce.

[26] Marchman and Bates (1994) report that past participle overgeneralisations (which are more frequent than syntactic overgeneralisations) represent only up to 17% of children's production.

3.8. EXPERIMENTAL METHODS

In the previous sections, it has been stated that Full Productivity (FP) is the result of a more abstract (fully schematic) representation, whereas Low-Scope Productivity (LSP) stems from a more concrete (lexically-based) one. Naturalistic data cannot (always) provide strong evidence to tell the two apart. Experimental methods represent a way around the issue as they allow a more systematic enquiry into children's knowledge. For instance, if a particular construction shows uneven productivity and rarely appears, experimental methods allow researchers to control for (and create) the pragmatic conditions for that construction to be used.

A high degree of lexical specificity in children's language does not rule out that they may be relying on fully-schematic patterns (see Dąbrowska & Lieven, 2005). At the same time, the fact that a small proportion (7%-21%) of children's spontaneous production cannot be accounted for in terms of lexically-specific units is not *per se* evidence that their language is not lexically-specific.

Clearly, there is a theoretical impasse: how can researchers disambiguate between LSP and FP?

The solution lies in trying to identify full productivity by ruling out the possibility that children could be relying on lexically-specific patterns.

3.8.1. IDENTIFYING FULL-PRODUCTIVITY

FP is clearly discernible when a speaker is capable of applying a rule or pattern to an item, even though the speaker herself/himself has never experienced that particular item in that particular pattern (or rule). That is, when the speaker cannot possibly have any lexically-specific pattern which could account for the sentence produced. Let us clarify this point with an example.

Mum, grandma and child (Rob) are in the living room. Grandma is not up to date with the kinds of cartoons Rob watches. One of these cartoons has a leitmotiv, which is a kind of greeting that involves the greeter kissing the greetee on his/her nose and then stroking his/her head. This is called *gorping*. Grandma has neither heard the word before, nor knows about such a form of greeting. The following conversation takes place:

6. 1 *CHI: grandma, I'll gorp you.
   2 *GRD: you what?
   3 *MOT: I <gorp> [!] you.
   4 *MOT: it is a kind of greeting that goes on in a cartoon
   5 Rob watches a lot.
   6 *CHI: 0 [=! gorps his grandmother and laughs].
   7 *MOT: grandma, what happened to you?
   8 *GRM: I got gorped by Rob.

Grandma hears *gorp* for the first time and immediately uses it in a morpho-syntactic environment (the passive) in which she has never heard or used it before, as she heard it only in transitive constructions (lines 1 and 3). This means that:

a) she can draw analogies between *I gorp you* and other transitive constructions by virtue of the fact that she can recognise that *I gorp you* is an instantiation of SVO or *AGENT-PROCESS-PATIENT* (*AGENT-PROCESS-PATIENT* → *I gorp you*) and so *I gorp you* activates such a fully schematic representation. Thus, she can transfer what she knows about other verbs to *gorp*[27].

b) her inventory of constructions is dense enough to draw connections between transitive and passive constructions:

   TRANSITIVE ⇢ PASSIVE.

c) she has a representation at least as abstract as *PATIENT-get-PROCESSed-by-AGENT* that she can activate in order to utter *I got gorped by Rob*[28].

Conversely, if her representation were lexically-specific, her other transitive schemas would be bound to specific verbs (e.g. *EATER-eat-THING-EATEN*). In this case, she would not be able to activate the fully schematic representation *AGENT-PROCESS-PATIENT*, resulting in her being unable to transfer what she

---

[27] Hence, she is capable of across-the-board generalisations based on fully-schematic representations.
[28] For clarity and easiness of exposition, I ignore that she might have a lexically-specific schema such as *I-got-PROCESSed-by-THING*, from which the sentences could be derived.

knows about other transitive verbs to *gorp.* The only experience she had with *gorp* is the schema *GORPER-gorp-GORPEE* and that is the only way she would be able to use the newly acquired verb[29].

3.8.2. USING NONCE LINGUISTIC MATERIAL

Controlling for children's input is then of crucial importance to gain an insight into their linguistic representation. If full productivity can be assumed when no possible lexically-specific pattern could account for speakers' sentences, it is necessary to elicit sentences that could **not** be derived from lexically-specific units previously acquired. Such a control is achievable by eliciting productivity with nonce linguistic material, with which no previous experience is possible. Many previous studies have tapped into children's linguistic productivity using **nonce verbs**.

An experiment that exposes children to real, existing verbs cannot disambiguate between LSP and FP, as adultlike behaviour with such verbs might be the by-product of children relying on well-rehearsed lexically-specific schemas. Nonce verbs allow researchers to control for the kind of input children receive. If children are presented with a nonce verb within a specific morpho-syntactic environment (e.g. the imperative *gorp her!*), researchers can be sure that this is the only occurrence of that verb encountered by participants. Children cannot have developed any lexically-specific pattern based on the nonce verb other than the pattern provided by the experimental stimulus. Hence, the extent to which they can use the nonce verb in a morpho-syntactic environment in which they have never encountered it (e.g. the passive construction *Peppa got gorped by*

[29] Note that, in this particular case, failure to use a passive sentence does not necessarily imply that the speaker does not have a more schematic representation of transitive constructions. What is needed in order to draw connections between transitive and passive constructions is an abstract representation of both of them. If the speaker has an *AGENT-PROCESS-PATIENT* representation, *I gorp you* could in fact activate the more abstract schema. However, if the speaker has not built a more schematic representation of the passive construction (and of its relationship with the transitive one), her knowledge of passive constructions is still lexically-specific (*KISSEE-get-kissed-by-KISSER*). In this case, the speaker cannot draw connections between the two constructions because the slots in her passive constructions are verb-specific; they are *KISSER* and *KISSEE*, not *AGENT* and *PATIENT*. As a result, she cannot activate a more abstract representation (*PATIENT-get-PROCESSed-by-AGENT*) to which she can connect her transitive schema. The result is that she cannot transfer what she knows about, say, *KISSEE-get kissed-by-KISSER* to *gorp*.

Emily) is a function of their ability to activate a schematic representation of the construction elicited.

3.8.3. EXPERIMENTAL EVIDENCE

Akhtar and Tomasello (1997) adopted a methodology that could be called "missing argument design". Their study 1 exposed ten children aged 2;9 to 3;8 to four nonce_verb-transitive_action pairs which missed either or both agent and patient. Each verb was presented in one of the following conditions:

a) Both arguments: *Cookie Monster is keefing Ernie*.

b) Agent only: *Cookie Monster is keefing*.

c) Patient only: *keefing Ernie.*

d) No arguments: *keefing*.

When transitive sentences were elicited, children were not able to correctly combine a verb with an NP argument which had not been heard in the stimulus. Thus, they were not able to combine the verb with a post-verbal patient NP if the stimulus had been presented with either conditions b (*Cookie Monster is keefing*) or d *(keefing*). Similarly, they were not able to combine the verb with a pre-verbal agent NP when the stimulus had been presented with either conditions c (*keefing Ernie*) or d (*keefing*). Four children never combined the verb with an argument that was not present in the stimuli. One child produced eleven correct NP arguments, even though she had never heard them occurring with the novel verbs. Five children produced a total of eleven attempts at combining a nonce verb with an NP argument with wich it had never been encountered in their experimental input: none of them succeeded. Of these eleven failed attempts, six agent NPs were uttered in post-verbal position and five patient NPs in pre-verbal position. Completely different results were yielded when children heard both arguments (*Cookie Monster is keefing Ernie*): 196 correct markings versus fourteen incorrect ones. Fig. 3.4 (overleaf) shows that when the newly acquired verb enters the system, it does so as an island of constructional organisation, independent from other constructions. If the stimulus exposes children to a *VO* order (fig. 3.4a), this is the way *keefing* enters the system and hence this represents children's knowledge of it. When *keefing* enters the system with *SVO* order (fig. 3.4b), children are conditioned to acquire a *KEEFER-keef-KEFFEE* schema and hence they will use the verb in an adultlike manner.

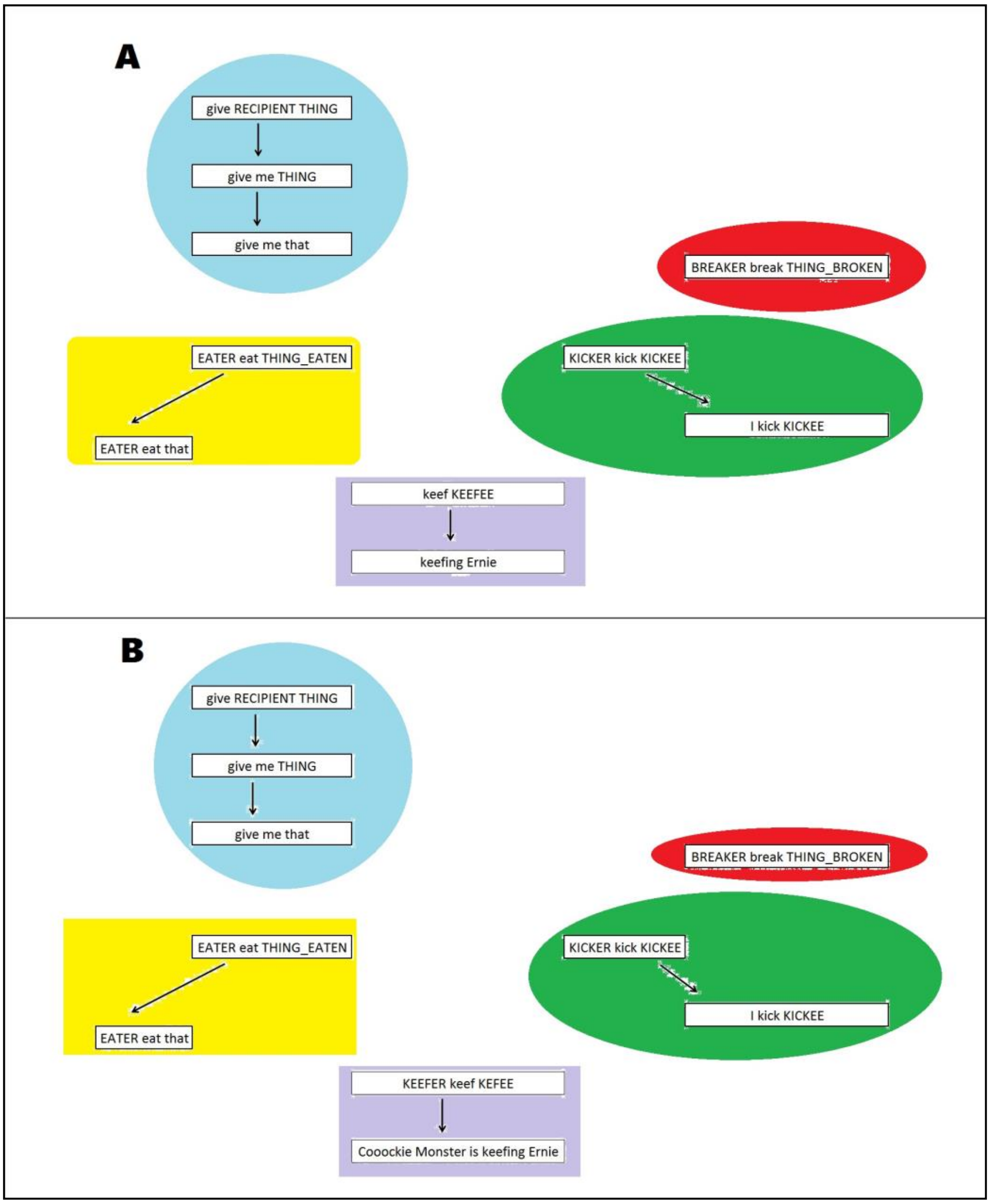


Figure 3.4: the verb *keef* enters children's lexically-specific inventories under VO (a) and SVO (b) conditions. Symbolic units are enclosed in rectangles in which *small letters* indicate lexically-specific material and *CAPITAL LETTERS* indicate slots. Arrows indicate relationships of elaboration/instantiation.

The overall picture shows that children's competence can be characterised as bound to knowledge of lexically-specific patterns. When a nonce verb was presented with V-NP sequences they used V-NP sequences with it, irrespective of the role (agent vs. patient) taken by the NP. That is, they used the lexically-specific patterns the experimental input provided (*keefing-THING*) and could not (fully) generalise what they knew about real verbs (*KICKER-kick-KICKEEE*) to newly encountered ones.

The main problem with Akhtar and Tomasello's (1997) design is that their stimuli were clearly a bit odd (Lewis, 2009; Ambridge & Lieven, 2011). Indeed, conceptualising the meaning of a verb is intrinsically linked with conceptualizing the entities that are put in relation to each other by the verb itself, i.e. its arguments (see Dąbrowska, 2004). Children are used to hearing verbs in contexts where these are put in relation to their arguments; exposing children to one-word stimuli (such as condition d) takes such relation away and may hinder form-meaning mapping.

Although this might indeed have influenced children's behaviour, it cannot be the whole story. Five children attempted to combine the nonce verbs with arguments which had not been encountered in the experimental input and produced both OV and VS orders. This is significant in two ways. Firstly, it shows that children understood the relational meaning of the nonce verbs, as they put them in relation to their doers and doees. Secondly, they did it in an input-dependent, ungrammatical way, "sticking" to the input (e.g. *Keef-THING*). Results can be interpreted as evidence that children's early productivity is limited (i.e. low-scope).

Akhtar (1999) exposed three age-groups (group A= 2;1-3;1, group B= 3;2-3;11 and group C= 4;0-4;4) of children to nonce_verb-transitive_action pairs within both grammatical (7) and ungrammatical (8-9) argument frames and then elicited transitive sentences with those verbs.

7. SVO

   Elmo dacking the car

   Elmo dacked the car

8. SOV

   Elmo the car gopping

   Elmo the car gopped

9. VSO

   Tamming Elmo the car

   Tamed Elmo the car

It is important to point out that verbs were presented with all necessary arguments and hence they were put in relation to both their patients and agents. Consequently, the semantics of the nonce_verb-action pair was transparent. Moreover, agents were [+ANIMATE] and patients were [-ANIMATE]. This mirrors the input children receive, as objects in CDS are overwhelmingly [-ANIMATE] and subjects are [+ANIMATE] (Tomasello, 2003). Hence, likelihood of animacy further favoured a correct semantic interpretation as to the roles played by the actors involved.

Once children had been exposed to those three verbs (each one randomly assigned to a different word order condition), the experimenter elicited transitive usages of them (*what's happening/What happened?*). All children were able to utter SVO sentences when the nonce verb had been presented within the same frame. As for the SOV and VSO stimuli, four-year-olds (4;0–4;4) were more likely to correct the input to SVO (*Elmo gopped the car*) than to adopt the ungrammatical word order (*Elmo the car gopped*) with which the new verb had been presented to them ($p<.01$). The same was not true for group A (2;1-3;1) and group B (3;2-3;11) children who were as likely to use the nonce verb with the ungrammatical word order of the input (*Tiger the fork dacked*) as to correct it to SVO (*Tiger dacked the fork*). The same **children** who produced ungrammatical word orders following the input also attended a **control condition** that exposed them to **a familiar verb (*push*)** presented within an ungrammatical frame (*Elmo the car pushed*). All children were more likely to use *push* with an SVO order than with the ungrammatical frame of the input (paired t (10)=4.37, $p<.01$; Akhtar, 1999, p. 348). Hence, children who were willing to **use** ungrammatical word orders with nonce verbs were **not** willing to use those same orders with the familiar *push*.

Such a contrasting behaviour can be accounted for in terms of lexically-specific knowledge. Most likely, children had previously experienced *push* and, consequently, could build a grammatical schema *PUSHER-push-PUSHEE*.

When production was elicited, *PUSHER-push-PUSHEE* was much more entrenched than *PUSHER-PUSHEE-push*. The latter had been encountered for the first time during the experiment and could not be as entrenched. Since entrenchment is a key factor in determining likelihood of activation (Dąbrowska, 2004; Langacker, 2008), *PUSHER-push-PUSHEE* was more easily activated. Children had a very entrenched schema that allowed them to overcome the input and produce adultlike sentences with *push*. The nonce verb had necessarily been encountered for the first time during the study and no other lexically-specific schemas for that verb could have been available. Hence, the new nonce verbs "entered the system" as the input presented them (e.g. *TAMMER-TAMMEE-tam*). As a result, children were willing to follow the input on a lexically-specific basis and were unable to transfer (or impose) what they knew about other verbs (e.g. *push*) to the newly acquired ones. Older children (4;0-4;4), instead, had a more entrenched fully-schematic representation (*AGENT-PROCESS-PATIENT*) that they could retrieve to overcome the ungrammatical input.

Overall, results suggest that a fully-schematic representation of the transitive pattern (*AGENT-PROCESS-PATIENT*) is not fully acquired before the end of the fourth year of life.

3.8.4. GOING BEYOND THE INPUT

The whole picture presented by these studies is more problematic than it appears. In all studies children **did** go beyond the input and produced grammatical sentences **using pronouns**, even when the input presented them with ungrammatical sentences (be they sentences missing arguments or sentences presenting ungrammatical orders).

Nine children out of ten in Akhtar and Tomasello's (1997) study produced at least one fully transitive sentence using pronouns (*he is keefing him*), even when the input presented them with missing arguments (condition b, c, and d). Similarly, 54% of corrections of VSO and SOV to SVO in Akhtar's (1999) study involved pronouns. Thus, converging evidence suggests that children acquire schemas built around pronouns whose SLOTs can be paraphrased as *PROCESS* (*he-PROCESS-him*, *I'm-PROCESSing-it*). The children in those studies did not have the chance to build grammatical schemas based on those nonce verbs and could not fully transfer what they knew about other verbs to them because of the lexically-bound nature of their units. Hence, when they followed the input (the safe solution in everyday life), they produced ungrammatical sentences. However, by relying on pronoun-based schemas they managed to either provide information missing in the input (Akhtar & Tomasello, 1997) or to overcome ungrammatical orders (Akhtar, 1999). Furthermore, when the children in Akhtar's (1999) study used ungrammatical orders (VSO, SOV), they, for all intents and purposes, never (1 out of 195 sentences) used pronouns[30]. This suggests that such pronoun-based constructions are entrenched enough for children to overcome ungrammatical information (the experimental stimuli).

The picture that emerges can be summarised as follows:

a) **Children's competence is input-dependent**: children overwhelmingly followed the experimental models, irrespective of whether those were grammatical (SVO) or not (VSO, SOV, missing arguments).

b) **Children's competence is concrete**: children younger than three years of age are generally known to be able to utter both intransitive and transitive sentences. However, when production was elicited, they could not transfer what they knew about other verbs to the nonce verbs in order to overcome ungrammatical models. This is evidence that the constructions on which children draw in order to be productive are independent of each other.

c) **Children are capable of generalisations from the very beginning** in that they categorised *keefing* as an instantiation of the semantic category *PROCESS* (*PROCESS* → *keefing*).

   Such a categorisation allowed them to fill the *PROCESS* slot of pronoun-based schemas (*he-PROCESSed-it*).

d) **Children *do* go beyond the input**, but in a constrained way; they either follow an already formed schema (as it is the case of productivity with pronoun-based schemas) or struggle to be grammatical (as was the case

[30] They never said *he him meeked*.

of VS and OV sequences when verbs were encountered with VO and SV orders, respectively).

e) **Children rely on lexically-specific schemas in order to be productive**: when they managed to be both productive and grammatical, they overwhelmingly did so by relying on pronoun-based schemas.

3.9. TOWARDS FULLY-SCHEMATIC CONSTRUCTIONS

By the age of four, English-speaking children have acquired a schematic representation of the transitive construction on which they can rely to overcome ill-formed input and deliver adultlike sentences. However, a precise insight into how children develop such abstract templates is probably the most underspecified area of UBAs to LA (Tomasello, 2003). Indeed, the debate between nativist and non-nativist scholars has very much dictated the research agenda. Nativist researchers primarily focussed on showing that children's underlying representation is abstract, possibly since before the very onset of language. UB researchers primarily focussed on providing evidence that children's early language is not fully-fledged. As a result, how children exactly move from a lexically-specific grammar to a more adultlike one is still (mostly) based on theoretical hypotheses rather than on (substantial) empirical data (Ambridge & Lieven, 2011).

3.9.1. STRUCTURAL AND FUNCTIONAL ANALOGY

According to UB researchers, more abstract constructions are developed by applying the same form-function schematisations that underpin the development of partially-schematic units; children draw analogies across lexically-specific schemas (fig. 3.5)

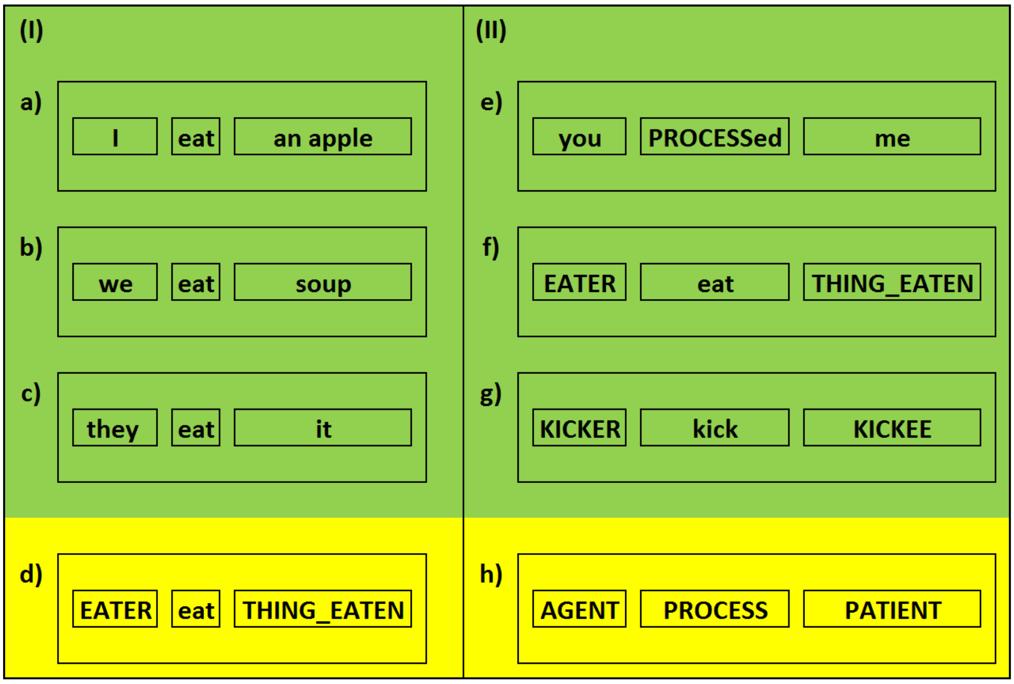


Figure 3.5 the development of the constructional schema *AGENT-PROCESS-PATIENT* (II), which is inferred by applying the same processeses of functional distributional analysis and structural alignment that underline the development of lexically-specific schemas (I).

Through schematisation, the entities analogised across are perceived as similar by abstracting away from their differences and retaining their common features. The lexically-specific schemas in fig. 3.5e-g and the fully schematic *AGENT-PROCESS-PATIENT* (fig. 3.5h) share both the pattern *NML-VERB-NML* and the fact that the pre-verbal element acts upon the post-verbal one. Since schemas inherit their meaning from their instantiations, *AGENT-PROCESS-PATIENT* inherits the meaning shared by fig. 3.5e-g (namely, X acts upon Y). Importantly, by abstracting away from the specificity of lexically-bound generalisations (*EATER*, *KICKER*), children develop superordinate structures that subsume those more specific units by means of relationships of elaboration: (*AGENT* → *EATER*; AGENT-*PROCESS-PATIENT* → *EATER-eat-THING_EATEN*).

In order to develop abstract templates, children have to align lexically-bound schemas as wholes and draw one-to-one connections among their different elements, the role these play, the relationships amongst them and, in English, their linear order. Thus, analogies across specific schemas must be made on the basis of both formal and functional alignment (Tomasello, 2003, 2006b).

In English, formal alignment means attending to word order. Tomasello (2006b) reports on Bauer (1996) who showed that even fourteen-month-olds are capable of imitating (hence learning) two-step and three-step actions. This implies that children have developed the non-linguistic skills to both partition a sequence into elements/steps and learn the order in which these must appear.

Even generalisations based on functional properties are attested from early on in development. For instance, two-year-olds overgeneralise the names of objects to novel objects that are perceived to perform similar functions (see Kelmer-Nelson, Russel, Duke & Jones, 2000).

Evidence from non-linguistic domains suggests that children are also able to draw analogies on the basis of the roles played by different elements within a unified scene. Tomasello (2003, 2006b) reports on a series of studies of Gentner and colleagues (Gentner & Markman, 1995, 1997; Gentner & Medina, 1998) which showed that children quite naturally attend to functional relationships. For example, children are shown picture A (a car towing a boat). They are then shown picture B, in which the **same car** as in picture A is being towed by a truck. Each child is then asked to identify which element in picture B is the same as the car in picture A. Children overwhelmingly choose the truck (a functional match) and not the perceptually identical car. Thus, children are naturally inclined to draw a connection between the truck and the car on the basis of their function in the scene, as opposed to their physical similarity.

3.9.2. THE RELATIONSHIP BETWEEN LEXICAL AND GRAMMATICAL DEVELOPMENT

An important theoretical assumption shared by various UBAs is the **lexicon-syntax continuum**. That is, there is no clear cut-off point between lexicon and grammar, as both single words and traditional rules are symbolic units, which differ “only” in specificity, generalisability, and complexity (Langacker, 2000, 2008).

The acquisition of grammatical patterns and lexical items are related in two ways:

a) In order to learn words, children have to single them out of concrete utterances. Singling out those words implies that at least some kind of rudimentary parsing is also going on.

b) When they learn lexical items (*tall*), children are also exposed to the distributional patterns associated with them (*tall_*).

Single-word utterances represent only 7% to 20% of CDS (Tomasello, 2003; Jusczyk, 1997). Consequently, the majority of words must necessarily be learnt by singling them out of multi-word utterances; children are capable of doing that well before (7th month) the onset of meaningful speech (Jusczyk, 1997). Singling a word out of an utterance implies that that utterance is (at least partially) being parsed/analysed.

For example, if a child who has learnt the word *mum* hears *mum is dancing*, s/he will be able to parse the sentence into *mum* and *is dancing*. Since the symbolic unit [[MUM]/[/mʌm/]] is known, the child may be able to map *is dancing* onto the action performed by her mother. The child has then acquired the complex symbolic unit in fig. 3.6.

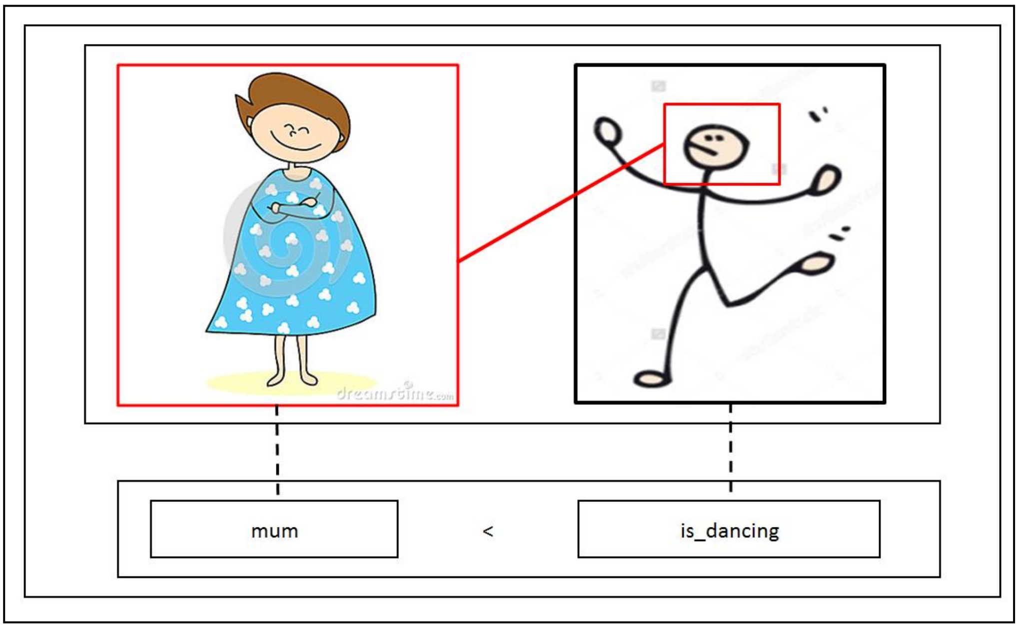


Figure 3.6: the symbolic unit *mum is dancing*. The phonological pole (bottom) is reported in standard spelling. The upper row, where the drawings are, represents the semantic pole. Dashed lines represent relationships of symbolisation. The red box surrounding the sketched man's head indicates that the meaning of *dancing* implies a dancing entity. The fact that the semantic pole of /mum/ is in red and is linked by a red line to the man's head (enclosed in a red box) indicates that *mum* is the dancing entity. The symbol "<" indicates linear order at the phonological pole.

The child might then encounter the sentence *the racoon is dancing*, extrapolate the unit *the racoon* from that sentence and map it onto its referent. This is facilitated by the fact that s/he has previously mapped *is dancing* onto the act of dancing. *The racoon is dancing* will help the child to learn the word *racoon* and strengthen the unit *is dancing.* Importantly, the child has now acquired a new symbolic unit (fig. 3.7)

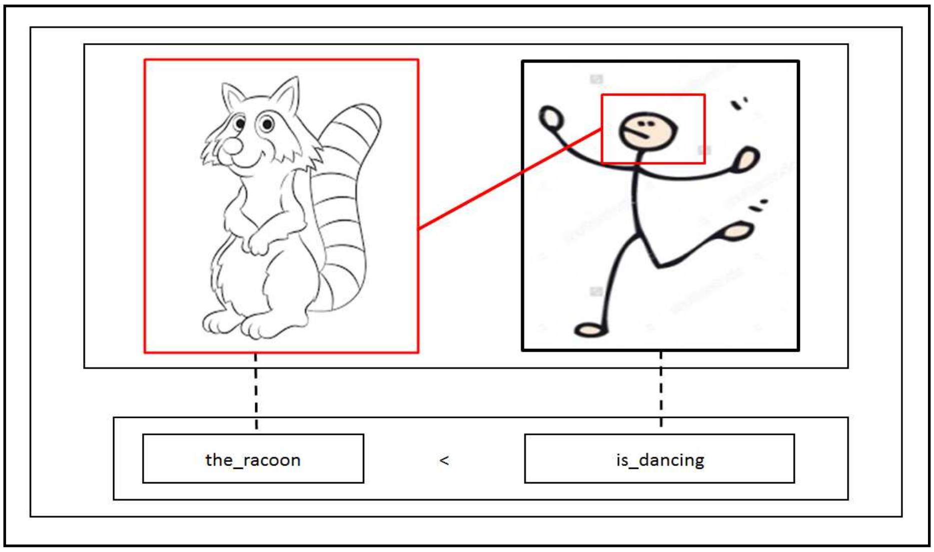


Figure 3.7: the symbolic unit *the racoon is dancing*. See the previous figure on how to interpret this figure.

The child could now draw analogies between *mum is dancing* and *the racoon is dancing* and infer the schema *DANCER-is-dancing* (fig. 3.8).

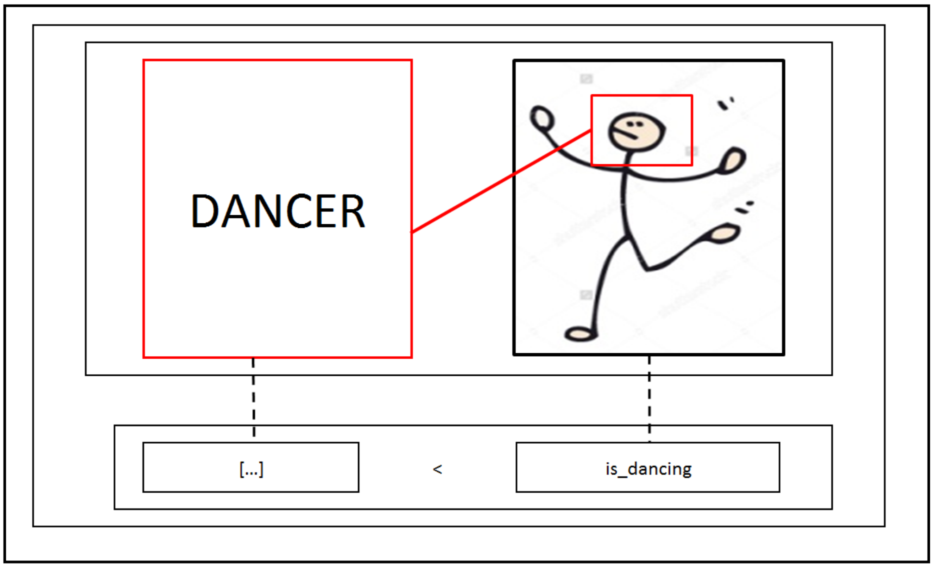


Figure 3.8: The schema *DANCER-is-dancing*. The interpretation of the figure is identical to fig. 3.6 and 3.7. What changes is that the dancing entity is specified only schematically at the semantic pole (*DANCER*) and not specified at all at the phonological pole ([…]).

Another important link between vocabulary growth and grammatical development is that when we learn words, particularly relational words, such as adjectives (*tall*), prepositions (*on*) and verbs (*hit*), we also store (and learn) their distributional patterns (Goldberg, 1995, 2006). Indeed, relational words (verbs, adjectives, prepositions) "make sense" only when put in relation to the entities participating in the relations they code. Learning the meaning of *hit* means learning that there are a *HITTER* and a *HITTEE* (see Lewis, 2009; Dąbrowska, 2009). Such knowledge amounts to knowledge of lexically-specific schemas (*tall_*, *_on_*, *_hit_*). Goldberg (1999) suggests that the development of more schematic patterns might be a necessity brought about by the memory overload that comes with vocabulary growth. In order to store more and more words (*ugly, bad*) and their associated frames (*bad_*, *ugly_*), categorising them into schematic patterns (QUALITY_) may indeed be necessary for an efficient "memory storage".

3.9.2.1. The Critical Mass Hypothesis

Bates and colleagues (Marchman & Bates, 1994; Bates & Goodman, 2001, Caselli, Casadio & Bates, 1999; Devescovi, Caselli, Marchione, Pasqualetti, Reilly & Bates, 2005) provide solid evidence that vocabulary size and grammatical complexity develop hand-in-hand. Bates and Goodman (2001) investigated the relationship between vocabulary size and grammatical complexity (the latter meant as a function of MLU) in twenty-seven children at ten, thirteen, twenty and twenty-eight months of age. They found that vocabulary at twenty months was a better predictor of MLU at twenty-eight months (r= +.83, p<.01) than MLU at twenty months (r. +.48, p<.05). Interestingly, different samples of MLU at twenty-eight months showed similar Pearson's values (r= +.75 to +.80). That is, vocabulary at twenty months and MLU at twenty-eight months had Pearson's values (r=+.83) that were virtually identical to those between separate samples of MLU at twenty-eight months and other separate samples of MLU at twenty-eight months (r=+70-80). The two authors point out that

> "no measure can correlate with another variable higher than it correlates with itself (i.e., Spearman's Law of reliability).
>
> (Bates & Goddmann, 2001, p. 138)

Thus, vocabulary at twenty months and MLU at twenty-eight months were statistically the same. The implication is that

> "One could be used as a stand-in for the other in predicting a child's rank within his/her group"
>
> (Bates & Goodman, 2001, p. 138).

Results were confirmed with a larger sample of children (1800) aged 2;4 to 2;6 and also held for atypical populations (William and Down syndrome, early focal lesions, late and early talkers).

Marchman and Bates (1994) found that morphological productivity (meant as a function of past participle overgeneralisations) was related to the number of verbs children knew (r +.56, $p<.0001$). They also noted that overgeneralisations sprang up after children had learnt about sixty/seventy verbs; the trend of the relationship was non-linear (quadratic term = $F(1,11) = 36.16$, $p<.0001$). A similar non-linear relationship was also found by Bates and Goodman (2001).

Those findings led to the hypothesis that children might have to encounter (and learn) a **critical mass** (a certain number) of exemplars (*HITTER-hit-HITTEE*, *EATER-eat-THING_EATEN*) before being able to extract general patterns (*AGENT-PROCESS-PATIENT*). This hypothesis is known as the **Critical Mass Hypothesis** (CMH, henceforth).

Devescovi et al. (2005) investigated the grammatical development of 233 Italian and 233 English-speaking children aged 1;6 to 2;6, using four measures of MLU. Results pertaining to both Italian and English showed that vocabulary was a strong predictor of all MLU measures. However, in Italian none of the measures showed a significant quadratic term. Conversely, in English all measures but one, showed a significant quadratic term. Thus, whereas the strong relationship between vocabulary growth and grammatical development appears to be attested cross-linguistically, the specific shape of their relationship may be language-specific.

3.9.2.2. Vocabulary is a better predictor than age

Bates and Goodman (2001), Caselli et al. (1999) and Devescovi et al. (2005) found that vocabulary is a better predictor of grammatical development than age. Devescovi et al. (2005) investigated each MLU measure through regression

(MLU ~ age + vocabulary + language) and found that vocabulary accounted for 12.7% to 20.9% of unique variance, depending on the MLU measure, whereas age accounted for only 0.7% to 2.9% of unique variance.

Findings are consistent with UBAs, which posit that vocabulary and grammatical development develop hand-in-hand and boost each other, as they are intrinsically related and underlined by the same processes of form-function mapping.

3.10. A WEAK TRANSITIVE SCHEMA

The evidence reviewed thus far indicates that English-speaking two-year-olds have a representation of the English transitive construction that is better describable as lexically-specific (*PUSHER-push-PUSHEE*), rather than fully-fledged (*AGENT-PROCESS*-PATIENT). Nevertheless, a more qualitative analysis of experimental evidence shows that young children have some more general competence of transitive patterns. Fisher (2002) notes that virtually all children in Akhtar's (1999) study (correctly) used nonce verbs presented within grammatical SVO patterns, whereas many of them (42% of two-year-olds and 25% of three-year-olds) avoided using nonce verbs encountered within ungrammatical frames. A further 42% (two-year-olds) and 17% (three-year-olds) of children produced both ungrammatical (*Elmo the car gopped*) and grammatical (*Elmo gopped the car*) sequences.

A later study of Noble, Rowland and Pine (2011) showed that English-speaking two-year-olds (2;2-2;10) are capable of interpreting reversible transitives correctly. Children watched two aligned screens: on screen A a duck acts upon a bunny and on screen B the bunny acts upon the duck. Children heard a transitive sentence describing either scene (*the duck is daxing the bunny*) and, when asked to point at the screen that matched the content of the sentence, they pointed to the matching screen (screen A) more often than it would have been expected by chance alone ($p<.04$).

Both Fisher (2002) and Akhtar (1999) interpret such results as evidence of children's early sensitivity to the patterns of English. Fisher (2002) also suggests that discrepancies between results on production (e.g. Akhtar, 1999) and comprehension (e.g. Noble et al., 2011) are explainable in terms of task difficulty (pointing towards a screen is definitely easier than producing a novel utterance) and semantic interpretation. She claims that children's linguistic knowledge should be investigated through comprehension studies, which, on the whole, suggest that children have some general understanding of the transitive pattern.

Tomasello and Abbot-Smith (2002) acknowledged such a pattern of findings and hypothesised that English-speaking two-year-olds have a weak representation of the transitive construction. Such a weakly entrenched schema can be exploited (retrieved) in certain comprehension studies (e.g. Noble et al., 2011), but it is not strong enough to support elicited production (e.g. Akhtar & Tomasello, 1997; Akhtar, 1999; Abbot-Smith, Lieven & Tomasello, 2001).

But what does having a weak representation of the transitive construction mean? How does this representation develop from weak to strong (or adultlike)? What kind of evidence is there to support Tomasello and Abbot-Smith's (2002) claim?

The first and third question can be addressed within the Adaptive Processing Approach (APA) model of children's knowledge (see Munakata, McClelland, Johnson & Siegler, 1997) (section 3.10.1). An insight into the second question (section 3.10.2) can be obtained within the prototype-based model of generalisation developed by Goldberg and colleagues (Goldberg, 1995, 1999, 2006; Goldberg, Casenhiser & Sethuraman, 2004; Boyd & Goldberg, 2012).

3.10.1. INFANT KNOWLEDGE AND GRADED REPRESENTATION

Munakata et al. (1997) investigated children's mastery of Object Permanence Concept (opc, henceforth) which the scholars explain as the ability to comprehend that an object continues to exist whether we perceive it or not and that it maintains its identity even if it changes location.

Opc is often investigated through *A-not-B* tasks (Piaget, 1952). Children witness the hiding of object X in location A and are then "sent" to find X. The same eight-month-olds who manage to retrieve X from A, often fail to retrieve X from a new hiding place B and go searching for X in the original location A. However, they are able to retrieve X when it is hidden in a location that makes it visible (e.g. under a transparent towel). This suggests that children are not fully capable of representing objects they do not see. However, studies adopting looking-time measures showed that infants as young as 3.5 months of age display competence of opc (they look longer at location B than location A).

Many researchers claim that ancillary factors, such as means-end deficit (i.e. inability to co-ordinate the removal of an obstacle object – e.g. a towel that is obscuring a task target-object, such as a toy train), prevent children from showing their knowledge of opc in *A-not-B* tasks. Munakata et al. (1997) take a different stand on such a line of results, claiming that children's task-dependence performance, rather than being a function of the development of ancillary factors, is grounded in the nature of children's knowledge, which is a matter of degree, rather than a presence/absence dichotomy. Specifically, they claim that different behaviours (experimental tasks) may require different degrees of representational strength of the knowledge being investigated.

According to Munakata and colleagues, knowledge is embedded in processing systems that guide overt behaviour. Such processing systems depend on the activation of specific processing units, which in turn depends on having to perform some specific tasks (e.g. engaging in a specific cognitive activity). The activation of such processing units is a matter of degree as it depends on the strength of the connections linking them. As children gain experience with a particular phenomenon, the connections amongst units underlying its representation are strengthened, yielding a more entrenched representation. The gradual development of an ever stronger representation of specific phenomena (be it the transitive construction or opc) enables children to increasingly demonstrate this acquired knowledge in a wider range of tasks that tap into the representations of these phenomena.

As a way of testing their hypothesis, the scholars built a simulation model which had an **internal representational layer** that represented the visual stimuli it experienced through a connected **input/output layer.** The two layers were connected through patterns of activation of processing units and their connection weights. Both the degree to which units were activated, and the strength of their connections had a graded nature (hence the connection between the two layers also had a graded nature).

The system was trained with visual stimuli involving a barrier and an object (e.g. a ball). As the ball stood still, the barrier moved towards it (from left to right, step 1-3) up to the point where the barrier occluded/hid the ball completely (step 4). After a variable length of time during which the ball remained occluded (e.g. steps 4-5), the barrier started moving back towards the left (step 6) up to a point where the ball was visible again (step 7). As the barrier started moving once more towards the left (step 6), the system could learn to anticipate the reappearance of the ball in step 7. Correct predictions were a function of the extent to which the network could maintain the processing units representing the visible ball (step 1 to 3) activated during the occluded steps (4 to 6). That is, the extent to which the representation of the occluded ball was similar to the representation of the visible ball (the network knowledge of the existence of the object when not visible, the network's opc).

The scholars ran twenty simulations in which each model underwent 1000 epochs of training. Results consistently indicated that

a) Even at early stages, some weak activation of the units representing the visible ball was attested during the occluded steps. Hence, the system could build a very weak representation of the ball very early on in the learning process.

b) As the system received more training (more experience), it was ever more able to keep activated the units representing the ball in a way that was similar across visible and occluded steps (i.e. its representation/knowledge of the object gradually strengthened).

A second model was then built in order to investigate the unique contributions of the representational system and of reaching skills to performance in *A-not-B* tasks. This second model was identical to the one previously discussed, but the **representational layer** (which represented children's developing representation of opc) was also linked through connection weights to an **output-reaching-system,** which represented children's ancillary skills, such as reaching skills (i.e. it represented the outcome of children's reaching behaviour). The aforementioned connections between the **representational layer** and the **output-reaching-system** transformed internal representations into outputs of the output-reaching-layer (i.e. into reaching behaviour needed to succeed in *A-not-B* tasks). In each of the twenty simulations, the researchers allowed the model to continue learning until it reached a developmental point called *early competence point* (**ecp**). A model which reached its ecp had a system responsible for predictions (the **input/output layer**) that was sensitive to both visible and hidden objects, whereas its **output-reaching-layer** was sensitive to visible, but not hidden, objects. Thus, the ecp simulated eight-month-olds' developmental stage, where children can retrieve visible objects but not occluded ones (which corresponds to the sensitivity of the **output-reaching-layer**) and look longer at hiding location B (which corresponds to the sensitivity of **the input/output layer**). Once a model reached its ecp, three types of simulations were run. In simulation A, **both the representational and output-reaching systems** were allowed to learn and develop. In simulation B, the **output-reaching-system** stopped learning and developing, whereas the **representational system** continued learning and developing. In simulation C, the inverse happened, the **representational system** was frozen and the **output/reaching systems** continued learning and developing. Results from simulation B showed that the model could improve its reaching behaviour as a result of the strengthening of the representational system alone. Importantly, the model's performance was similar across simulation A (the complete model) and B. This was not true for simulation C, where improvements in reaching abilities alone yielded only small (and inconsistent) improvements.

The results lead us to conclude that:

a) the ability to represent objects even when not visible, that is, knowledge of opc (which is representative of children's knowledge), can be characterised in terms of a representation that lies on a continuum from weak to strong;

b) experience induces learning that allows the system to strengthen its representation;

c) changes in representational strength alone can improve performance in specific tasks;

d) looking tasks can be carried out successfully with a weaker representation than that required to successfully perform in reaching tasks. Hence, different tasks require different degrees of representational strength.

If children's knowledge can be described as involving representations that grow from weak to strong as a function of experience, it is not implausible to posit that children's knowledge of the transitive schema will also grow as function of experience. Fig. 3.9 shows the proportion of English-speaking children who produce adultlike transitive utterances in experimental settings.

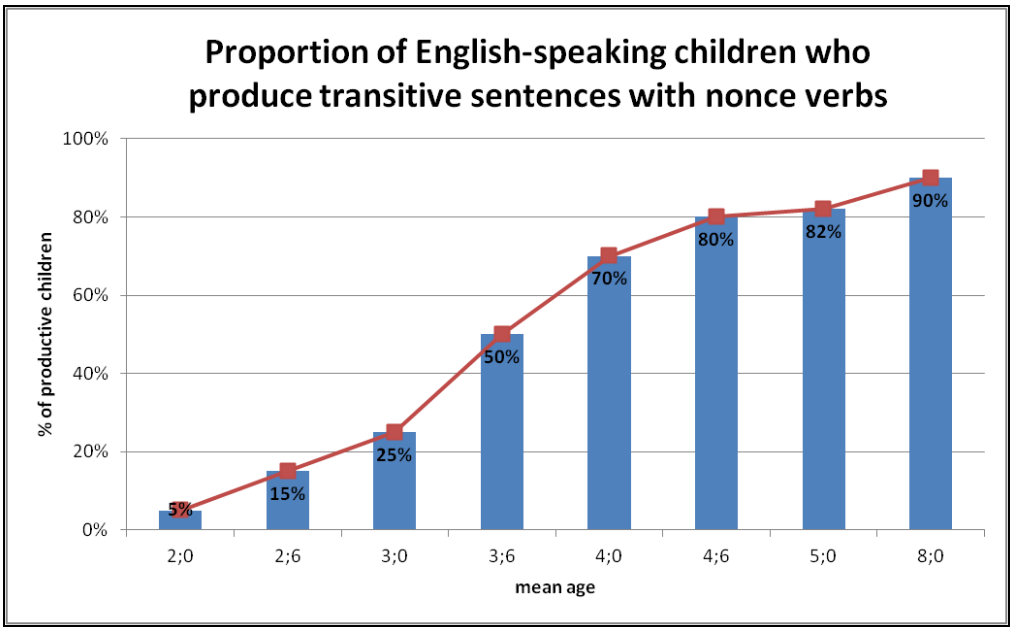


Figure 3.9: Proportion of English-speaking children who produce adultlike transitive sentences with nonce verbs. Based on Tomasello (2006b, fig. 6.1, p. 266).

The pattern that emerges shows a gradual development that can be thought of as a function of the piecemeal entrenchment (strengthening) of a schematic representation of the transitive construction[31]. English-speaking children start developing a weak representation of the transitive construction early on in development (during the third year of life). This weak representation allows them to perform well in comprehension (Noble et al., 2011), but not in production (Akhtar & Tomasello, 1997; Akhtar, 1999; Abbot-Smith et al., 2001) studies. As the representational strength of the transitive construction increases as a function of linguistic experience, children become able to bring such knowledge to a wider variety of tasks[32].

Cross-linguistic evidence is consistent with the hypothesis that task difficulties cannot explain away the poor performance in production studies of young English-speaking children.

Dąbrowska and Tomasello (2008) trained two (2;3-2;9) and three (3;0-3;3) year-old Polish-speaking children with nonce verbs governing instrumental case, combined with either masculine (*MEEKER+mikuje+MEEKED-**em*** "MEEKER+meeks+MEEKED-**instr.m.sg**") or feminine (*MEEKER+mikuje+MEEKED-**a*** "MEEKER+meeks+MEEKED-**instr.f.sg**") nouns. They then tested children's ability to combine verbs they encountered only in combination with nouns of one gender (INSTR.M.SG) with nouns belonging to the other gender (INSTR.F.SG). If children had a schematic representation of the instrumental case, they should be capable of combining a nonce verb with both masculine and feminine nouns, irrespective of the specific

[31] Note that figure 3.9 could be given two interpretations. On the one hand, it might be argued that whether children are productive or not is a dichotomous distinction (productive vs. non-productive) and the fact that a minority of children show productivity at the age of two years, whereas most children show productivity later on in development (around the age of four years), can be interpreted in terms of individual differences (see Dąbrowska (2004) for a brief summary of individual differences in acquisition). Some children simply become linguistically productive earlier than others. However, the graph can also be interpreted in terms of graded representation: at age two there is only a 5% chance that children will be productive and at age four a 70% chance. The fact that the youngest children in Akhtar's (1999) study behaved inconsistently seems to be consistent with the latter interpretation. However, the attested individual differences in acquisition and a view of children's linguistic knowledge, as graded in representational strength, are not mutually exclusive. Nevertheless, I shall focus on the second interpretation, as this sub-section aims at discussing how models of children's knowledge as graded in representational strength could potentially handle the developmental data that emerge from LA studies.

[32] As to how children's linguistic competence may develop from weak to strong, it is discussed in light of Goldberg's prototype-based model of generalisation in the next section (3.10.2), (see Goldberg, 1995, 1999, 2006; Goldberg et al., 2004; Boyd & Goldberg, 2012).

gender condition (MASC or FEM) with which they encountered it (Dąbrowska & Tomasello, 2008). About 39% of two-year-olds and 81% of three-year-olds successfully combined a nonce verb with an instrumental-inflected noun whose gender differed from the gender condition encountered during training. Hence, three-year-olds, and to a lesser extent two-year-olds, had a schematic representation of the instrumental case they could bring to the experimental task. Importantly, the proportion of productive Polish-speaking two-year-olds is much higher (39%) than the average proportion of their English-speaking peers who productively use nonce verbs in transitive constructions (5% to 25%; see Tomasello, 2000b, 2006b; refer back to fig. 3.9).

If the poor performance of English-speaking children were exclusively due to task difficulties, why wouldn't such difficulties affect Polish-speaking children's performance with the same task?

Bates and MacWhinney's (1987) **Competition Model** (CM, henceforth) provides an angle from which it is possible to approach such a question. According to the CM, speakers rely on different cues to map clausal elements onto their functions (e.g. their thematic roles). A cue can have different degrees of **reliability** and **availability**. The former is a function of how many times a particular morpho-syntactic phenomenon (e.g. accusative case, post-verbal position) maps onto a particular role (patient) as a proportion of its total occurrences. Cue availability is a function of its frequency. The product of availability x reliability determines a cue's **validity**. Cues can either converge or enter into competition to assign a thematic role to a specific clausal element. Importantly, cues have **processing costs**. **Local cues** score lower in processing cost than **topical or global cues**. A **local cue** operates at the word level, in that it can be processed with no or little consideration of other clausal elements. Case marking is a local cue as, in order to infer the role played by an element in a sentence, one only needs to attend to the specific inflection a word takes. A **global cue** is

> "a cue which spans two or more disparate and perhaps discontinuous elements"
>
> (Bates & MacWhinney, 1987, p. 180)

and therefore puts a heavier burden on the processing system (see also Abbot-Smith & Serratrice, 2013). Word order is a global cue, as in order to infer the role that *the cat* plays in *John killed the cat*, one has to process both verb and NP and their linear order. According to the CM then, word order should be more difficult to process and acquire than case marking.

Slobin and Bever (1982) exposed Turkish, Croatian, Italian and English-speaking children to sequences of one verb and two NPs (NVN, VNN, NNV) and then asked them to act out who did what to whom. Turkish-speaking 24-to-28-month-olds, who could attend to a local cue (case), interpreted the stimuli correctly 82% of the times. Conversely, only 58% (below the 67% chance level established by the authors) of the answers given by their English-speaking peers, who had to rely on a more topical cue (word order), gave NP-V-NP sequences an SVO interpretation (the only possible one in adult English).

It is then possible that the language-specific nature of the cues children must attend to determines the pace at which such cues are acquired and the pace at which children develop entrenched schematic representations of the grammatical phenomena expressed through such cues. Polish-speaking children, who can rely on local cues, develop a strong schematic competence earlier than their English-speaking peers. Task-specific difficulties do not hinder their performance because they can rely on fairly solid representations. Conversely, English-speaking children, who must attend to a more topical cue, are still developing their transitive construction. This combination of task difficulties and English-speaking children's weak schematic representation may be the reason behind their poor performance.

3.10.2. A PROTOTYPE-BASED MODEL OF GENERALISATION

Goldberg et al. (2004) propose a model of how generalisation (hence schematisation) happens. The main idea behind their proposal is that constructions are linguistic patterns whose meanings and functions (or at least part of them) are not fully inferable by the meanings of the elements that appear in them (Goldberg, 2006). In *Mark kisses Rob*, the meanings of *Mark*, *Rob* and *kisses* are not enough to infer that Mark is the kisser. The conventionalised transitive construction (*AGENT-PROCESS-PATIENT*) is a form-meaning pair in which the pre-verbal element has to be understood as the agent.

Goldberg et al. (2004, p. 293) note that

> "verbs that are closely related semantically do generally appear in the same Argument Structure Constructions" (**ASC**, henceforth).

For example, *move* (10b) and *put* (10a) share the fact that they often appear in the caused-motion construction (**cmc**, henceforth), that maps onto a scene in which *X causes Y to move in location Z.*

10. a) mum put your pacifier on the table.

    b) I moved the chair into the living room.

Goldberg and colleagues (Goldberg, 1999, 2006; Goldberg et al., 2004) also observe that the token instances of different ASCs tend to be dominated by a set of light verbs (*go*, *do*, *make*, *give* and *put*), which have a broad general semantics. Such a broad semantics makes the use of those verbs appropriate for a very wide range of situations and hence makes them very frequent. Goldberg et. al (2004) analysed three types of constructions[33] in the spontaneous speech of twenty-seven children aged twenty to twenty-eight months and fifteen mothers[34]. They found that each ASC tended to be dominated by a single (light) verb. For example, *put* accounted for 31% of all tokens of the cmc in the children's corpora and for 39% of all tokens of the same construction in the mothers' corpora. Hence the cmc is dominated by the light verb *put*, which can therefore be thought of as its **prototype**.

Goldberg (1999) hypothesises that learners are likely to draw correlations between the meaning of the prototype verb and the construction it dominates. As a consequence, the meaning of the verb is extended to the construction in such a way that, even when that particular verb is not present, its general semantics is retained by the constructional pattern itself (Goldberg, 1999).

Goldberg et. al (2004) exposed three groups of undergraduate students to a nonce construction indicating appearance (something/someone appeared on the

[33] The ditransitive, the intransitive and the caused-motion construction.
[34] From the Bates corpus (Bates, Bretherton & Synder, 1988).

scene; *AGENT-LOCATION-VERB*). The concrete instantiations of such a construction contained nonce verbs and real nouns (*the king the ball mopo-ed*). The IV was input, which had three levels: none, balanced and skewed. For each condition, a different group of participants received no training, balanced training and skewed training. There were five nonce_verb-action pairs. One of the verbs was designed as the prototype for the construction. Hence, its meaning was more general and more extendable to a general scene of appearance. Participants in the skewed input group were exposed to the prototypical light verb 50% of the times. Those in the balance group received an input more evenly partitioned. Subjects were then tested with a forced-choice comprehension task. Novel nonce verbs were presented within the same construction. For each novel verb, participants had to choose between two scenes: one which matched the meaning of the construction and one which did not. The skewed input group (M=5.1, sd= 0.96) significantly ($p<.01$) outperformed the balanced group (M=3.8, sd= 1.48).

Results indicate that the statistical skewing attested in language use (and in children's input) helps schematisation, as it helps the formation of a prototype to which learners can assimilate new elements on the basis of similar form-function mapping.

If Goldberg and colleagues are on the right track, it is possible that when children learn a new lexically-specific transitive pattern (*HURLER-hurl-HURLEE*) they do so by apprehending it as an extension vis-à-vis a perceived prototype. According to Goldberg (1999), the prototype of the transitive construction is *DOER-do-DOEE*. Hence, a new transitive verb may initially be apprehended as an extension vis-à-vis *DOER-do-DOEE:*

11. *DOER-do-DOEE* ---► *HURLER-hurl-HURLEE*.

By using *DOER-do-DOEE* to categorise *HURLER-hurl-HURLEE*, one has to filter out the two constructions' conflicting specifications and conceive a superordinate structure that is instantiated by both schemas (namely, *AGENT-PROCESS-PATIENT*). A unit enters the system through entrenchment. When *HURLER-hurl-HURLEE* is apprehended as an extension vis-à-vis *DOER-do-DOEE*, the schematisation (*AGENT-PROCESS-PATIENT*) will leave a feeble trace in children's minds. When another lexically-specific pattern (e.g. *DESTROYER-destroy-DESTROYEE*) is acquired by apprehending it as an extension vis-à-vis the prototype, *AGENT-PROCESS-PATIENT* will become a bit more entrenched, and so on[35].

At this point a weak representation of the transitive schema is being formed. However, such a weakly entrenched schema can only be evoked via the activation of more concrete and entrenched units (e.g. *DOER-do-DOEE*). As children acquire more and more lexically-specific transitive patterns which are categorised as extensions vis-à-vis the prototype, the superordinate structure (*AGENT-PROCESS-PATIENT*) is evoked more and more frequently. As this happens, children's schematic representation of *AGENT-PROCESS-PATIENT* strengthens, up to the point at which (around four years of age) it attains status of unit and becomes available to sanction new expressions (*AGENT-PROCESS-PATIENT* → *GORPER-gorp-GORPEE*)[36].

At this point, the network metaphor that has so well served the discussion thus far shows its flaws. It has been previously stated that children's inventories have much poorer connections, whereas adults' inventories are highly interconnected. It is probably better to say that, in children's inventories, those relationships are weaker, less entrenched, because they are based on much fewer and less entrenched units. Relationships across units are not dashed and continuous arrows and lines that link boxes, but cognitive processes that reside in patterns of neurological processing (Langacker, 2008), and so is the degree of entrenchment of a unit.

---

[35] Such a hypothesis is consistent with the strong relationship between vocabulary growth and linguistic development discussed in section 3.9.2. If learning relational words (e.g. *kick*) is inseparable from learning their distributional properties (*KICKER-kick-KICKEE*) and hence inseparable from learning lexically-specific schemas, any time that a new relational word is learnt, a lexically-specific schema is apprehended as an extension vis-à-vis a prototype. As children acquire more and more words (hence schemas), they more and more frequently evoke a superordinate structure A', necessary to apprehend a new expression B as an extension vis-à-vis a prototype A. The more words (schemas) children learn, the more entrenched the superordinate structure A' instantiated by the lexically-specific patterns (A and B) linked to those words will become.

[36] Hence successful performance in comprehension studies might be interpreted as children apprehending the new patterns (*The duck is daxing the bunny*) as extensions vis-à-vis a prototype. Production may be a more demanding process because, in order to produce the target sentence, children may have to sanction the new expression (*Elmo keefed the car*) as an instantiation (or elaboration) of a fully-schematic unit (*AGENT-PROCESS-PATIENT*).

3.10.3. THE ROLE OF ENTRENCHMENT

Claiming that children have a weak transitive schema somehow implies that they have at least weak knowledge of such a constructional pattern. Fisher (2002) seems to argue that such a weak representation constitutes the kind of innate knowledge with which children are endowed.

However, an early weak representation can satisfactorily be accounted for without resorting to endowed knowledge. Firstly, Munakata et al.'s (1997) model could form a weak representation of opc very early on in development on the basis of experience, without having any pre-constructed knowledge of it. Linguistically, this may mean that all is needed is the ability to learn a frequent prototype and some analogy skills through which new patterns could be apprehended (and learnt) as extensions vis-à-vis that lexically-specific prototype. Hence, ability to learn from the input is all that is needed in order to create such an early sensitivity. Secondly, such an early sensitivity is accountable for by factoring in the role of entrenchment (and therefore frequency). Entrenchment, which is a function of token frequency, determines the representational strength of a given unit: the more an expression is used and/or heard, the more it entrenches[37]. Hence, all things being equal, **more frequent expressions will be acquired (and schematised) earlier than less frequent ones**.

A cross-linguistic look at production of passive sentences represents a case in point. Spontaneous production of full passives by English-speaking children is normally attested only around the fifth year of life (Tomasello, 2003; Brooks & Tomasello, 1999). In contrast, children speaking non-European languages (Maya, Inuktitut, Sesotho) produce and overgeneralise passives as early as during the third year of life (Brook & Tomasello, 1999). UB researchers account for such a cross-linguistic discrepancy in light of the frequency with which passives occur in children's input; they are incredibly infrequent in the CDS of young English-speaking pre-schoolers, whereas they are very frequent in those non-European languages.

---

[37] In a similar fashion, Munakata et al.'s (1999) model strengthened the connections symbolising its representation of the ball as it received more and more training.

Brooks and Tomasello (1999) exposed two age-groups (A=2;7-3;0 and B=3;3-4;2) of English-speaking children to two experimental conditions with two nonce verbs: intensive training with transitive frames (*Big Bird is meeking the car*) and intensive training with passive frames (*the car is getting meeked by Big Bird*). They then elicited production of full passive sentences. About 90% of the children who were trained with passives produced passive structures during the test phase, whereas only two children (12%) who had been trained with transitives did so ($F(1,52) = 53.30$, $p<.001$; Brooks & Tomasello, 1999, p. 32). Hence, when the frequency and amount of passives to which English-speaking children are exposed are experimentally manipulated, even two-year-olds (2;7-3;0) are capable of producing full passives, which is about one year earlier that the literature normally attests. This suggests that the late acquisition of passives by English-speaking children is a by-product of the low frequency of the construction in the input language. Similarly, early sensitivity to the transitive pattern is likely to be the by-product of its high frequency (up to 41% of Anglophone CDS includes some kind of *AGENT-PROCESS-PATIENT* pattern; see Tomasello, 2003)

Thus, different constructions (passives, transitives, intransitives) are acquired at different pace and at different developmental points and such differences appear to be functions of their frequency in the ambient language. Furthermore, since different languages exploit some constructions more than others, children acquiring different languages learn the same constructions at different developmental points.

3.11. SUMMARY

LA is about learning and developing a network of symbolic units (constructions) – i.e. form-meaning pairings varying in size and degree of specificity - and the connections that link them. Children learn those units by abstracting them from the concrete strings they encounter. Initially, complex units are learnt as single big words (*I_eat_soup*), in which a string of phonemes maps onto a (holistic) meaning (THE_SPEAKER_EATS_SOUP). As children learn similar strings (*we eat everything*, *you eat pasta)* they can start drawing analogies across them and develop lexically-specific schemas (*EATER-eat-THING_EATEN*). Initially, children's linguistic competence can be described as mastery of an inventory of lexically-specific units, which are either fully lexically-specific (*you_eat_it*) or only partially schematic (*EATER-eat-THING_EATEN*). Experimental studies indicate that each lexically-bound schema represents an island of lexically-specific knowledge mapped onto a specific meaning, with little or no connection to other islands (e.g. children cannot transfer what they know about *push* to newly acquired nonce verbs; see Akhtar, 1999). Thus, their linguistic productivity appears to have a low-scope nature, as it is bound to lexically-specific generalisations.

As children learn more and more lexically-specific constructions (*KICKER-kick-KICKEE*, *KISSER-kiss-KISSEE*), their ability to draw analogies improves. More schematic and adultlike patterns (*AGENT-PROCESS-PATIENT*) are learnt by drawing generalisations from the lexically-specific schemas children inferred from the input. The acquisition of such fully-schematic patterns is, however, a slow and continuous process and grammatical development goes hand-in-hand with vocabulary growth. English-speaking children start showing some general understanding of the schema *AGENT-PROCESS-PATIENT* half-way through the third year of life. However, such knowledge reaches adultlike competence only during the fourth/fifth year of life.

Page intentionally lef blank

4.

DESIGN RATIONALE

Converging evidence from both naturalistic (Dąbrowska & Lieven, 2005; Lieven et al., 2009) and experimental (Lewis, 2009; Akhtar & Tomasello, 1997; Akhtar, 1999) studies is consistent with an account of children's early language as concrete (i.e. lexically-bound). Later on in development, fully schematic templates are abstracted in a piecemeal fashion (Tomasello, 2003) and their representational strength grows gradually, from weak to strong, as a function of cognitive maturation and linguistic experience.

However, three main pending issues haunt UB scholars. Firstly, since UB research has overwhelmingly focussed on providing evidence of the lexically-specific nature of early language, accounts of how children exactly develop fully-schematic units are still (mostly) based on theoretically grounded hypotheses, rather than on (solid) research evidence (but see Childers & Tomasello, 2001; Goldberg et al., 2004; Boyd & Goldberg, 2012).

Secondly, UB researchers have overwhelmingly focussed on the acquisition of English (Tomasello, 2003), which is a rather atypical language; its words are mostly monosyllabic, its morphological system is highly impoverished and it presents a unique rigidity of, and reliance on, word order (Bates & MacWhinney, 1987). Thus, the extent to which results regarding English-speaking children would generalise to other languages, particularly typologically different ones, needs much further cross-linguistic research.

Finally, both nativist and UB researchers have still much work to do in order to account for how children constrain their generalisations (for a review, see Tomasello, 2006b; Ambridge & Lieven, 2011). For example, once children have abstracted a pattern like *TRANSFER-RECIPIENT-THING*, how do they learn that some verbs (*bake Mark a cake*), but not others (**Explain me that*), can instantiate that pattern?

This research aims to gain an insight into the second question and investigates the extent to which a UBA can account for the acquisition of Italian, and hence can be said to have cross-linguistic validity.

4.1. THE QUEST FOR CROSS-LINGUISTC VALIDITY

Since Tomasello's (2000a, 2003) first calls for cross-linguistic evidence, a growing body of UB research has investigated the acquisition of languages other than English, such as German (Abbot-Smith & Behrens, 2006), Spanish (Aguado-Orea, 2004), French (Matthews, Lieven, Theakston & Tomasello, 2007), Portuguese (Rubino & Pine, 1998), Polish (Dąbrowska & Tomasello 2008; Dąbrowska, 2008) and Cantonese (Chan, Lieven & Tomasello, 2009), to name but a few.

This research posits itself within this quest for cross-linguistic evidence and investigates the development of morpho-syntactic competence in Italian-speaking children. Italian is a Romance language whose highly inflected morphological system is very rich, whose words are mostly tri-syllabic and whose word order is flexible and determined by discourse pragmatics. Hence, it differs from English in many morpho-syntactic respects. Importantly, UB research on Romance languages (Spanish, Portuguese) has mostly focussed on morphological development and mostly used naturalistic methods (e.g. Aguado-Orea, 2004). By investigating both morphological and syntactic development and by adopting both naturalistic and experimental methods, this study aims to provide a more precise insight into the acquisition of Romance languages.

4.2. DESIGN AND RESEARCH QUESTIONS

Previous research on Italian-speaking children (Pizzuto & Caselli, 1992, 1993, 1994; D'Odorico et al., 2011) is consistent with UB models, as it shows that children's linguistic productivity and development are gradual and uneven. Although such outcomes are highly informative, uneven distribution of lexical items across morpho-syntactic patterns is often a property of language *per se* and it therefore does not constitute conclusive evidence of a lexically-specific competence (Yang, 2009).

In order to provide evidence that a UBA can account for the acquisition of Italian, it must be proven that:

a) the linguistic production of Italian-speaking children is describable in terms of lexically-specific units that can be characterised as low-scope generalisations of the concrete strings that children themselves have previously encountered.

b) Italian-speaking children's linguistic productivity provides little (or no) evidence of an adultlike productivity. That is, children's productivity is better characterised as low-scope, rather than as fully-fledged.

Thus, whether a UBA can account for the acquisition of Italian is investigated by means of two research questions:

a) Can Italian-speaking children's early language be accounted for in terms of lexically-specific units acquired from the concrete language that children themselves have previously experienced?

b) To what extent can Italian-speaking children be said to rely on (have mastered) fully-schematic constructions/patterns?

4.2.1. NATURALISTIC INVESTIGATION

Research question (a) is answered by means of a longitudinal study (Part II of this work) that adopts the traceback method discussed in chapter 3.5. Such a method allows a fairly psychologically realistic insight into children's language because, rather than focussing on the distributional properties of morpho-syntactic patterns, the analysis tries to account for children's language in terms of putative lexically-specific units attested in their own linguistic experience[38]. Since lexically-specific schemas appear around the age of 2 (Tomasello, 2006a-b), the naturalistic study analyses the language spoken by an Italian-speaking two-year-old.

4.2.2. EXPERIMENTAL STUDY

As previously observed, the high degree of lexical specificity that emerges from naturalistic studies is not *per se* conclusive evidence that children do not rely on more abstract patterns. Indeed, such results might also stem from the interaction of some statistical properties of language (Yang, 2009), children's limited vocabulary (Lieven et al., 2003) and the routinised situations they experience

[38] Importantly, it is not claimed that, when derivations are successful, these are the exact ways in which children "assemble" their sentences or that the units identified are the exact units on which they rely. Rather, the method aims to provide evidence that children's production can *potentially* be accounted for in terms of lexically-specific units.

(refer back to chapter 3.7). Experimental methods allow researchers to investigate children's language right from the other end of the "schematicity continuum", as they allow manipulations (controlling for pragmatic situations, for children's input, and so forth) which can help to gain an insight into the extent to which specific aspects of early language are fully-schematic (refer back to 3.8.1 and 3.8.2).

Thus, research question (b) is investigated by an experimental study (Part III of this work), which taps into children's morpho-syntactic productivity with both a nonce verb and a familiar verb. The assumption is that the extent to which children can use a nonce verb in a morpho-syntactic environment (e.g. the transitive construction) in which it has not been experienced before, is the extent to which their representation of that specific morpho-syntactic phenomena (the transitive construction) is fully-schematic.

The design of the experiment is modelled on Lewis' (2009) and Akhtar and Tomasello's (1997) studies. Lewis (2009) exposed children to imperative sentences and elicited production of declarative transitive sentences. As will be discussed in detail in chapters 10 and 11, when the language at issue is Italian, using imperative stimuli while prompting declarative sentences allows the elicitation of both morphological and syntactic productivity by asking children to perform one task only (hence lightening their cognitive effort). For instance, in order to transform an imperative experimental stimulus like (1) into a declarative, past (*passato prossimo*) transitive sentence like (2), children must change both the verb's morphological ending (*mangi-**a=l-o*** "eat-IMP.2.SG=3.ACC-M.SG" becomes *mangi-**a-t-o*** "eat-TV(conj.I)-PTCP-M.SG") and the stimulus' word order (from Vo to oV; see chapter 11). This would be evidence of morpho-syntactic productivity.

1. *Mangialo!*
   mangi-a=l-o!
   eat-IMP.2.SG=clitic.3.ACC-M.SG
   "Eat it/him!"

2. *L' ha mangiato*

l'	*ha*	*mangi-a-t-o*
clitic.3.ACC	has	eat-TV(conj.I)-PTCP-M.SG

   "(She/he) ate it/him."

Children start building lexically-specific schemas roughly around the age of two, and by the time they are four, their language has so significantly improved that, for example, they have a schematic representation of the transitive construction (Tomasello, 2003; Akhtar, 1999). Hence, participants are children aged 2;2 to 5;0, divided into three age-groups (two-year-olds, three-year-olds and four-year-olds), plus one control group of adults. Comparative analysis of how each age group performs can provide an insight into how linguistic representation develops over time.

4.3. A FINAL REMARK

The rationale behind the current design is that this twofold methodology may provide evidence that could reasonably lead to drawing fairly solid conclusions as to the nature of early Italian. For instance, if the naturalistic study showed that the majority of the child's spontaneous production could be accounted for in terms of lexically-specific units, and the experimental study showed that his peers' linguistic productivity could not be described as fully-schematic (not in an adultlike way, at least), then evidence would converge towards a lexically-specific account of children's language.

At that point, claiming that naturalistic results are a by-product of external and ancillary factors (such as children's vocabulary, natural distributional properties of linguistic patterns, etc.) and that children's experimental performance is a by-product of the development of ancillary factors, would run against converging evidence.

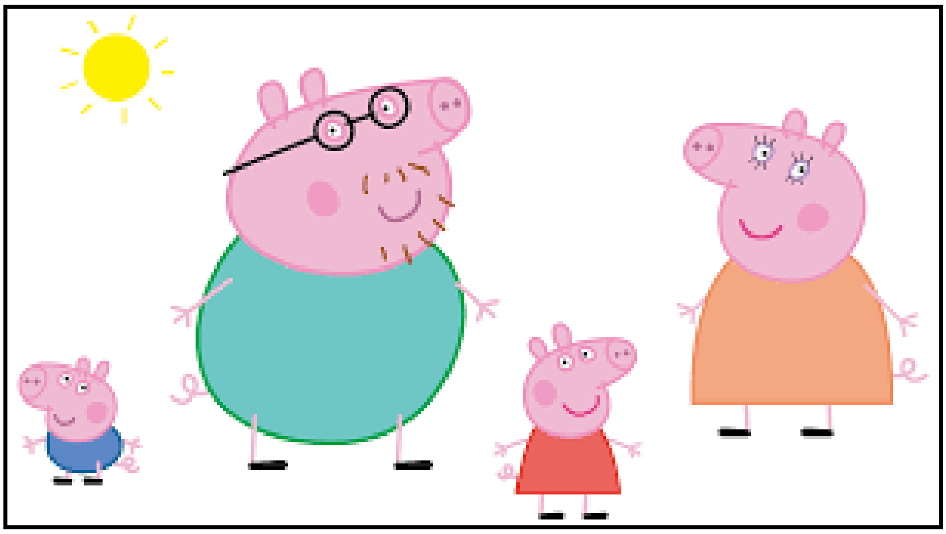
More specifically, claiming that the results of the longitudinal study are the by-product of the child's limited vocabulary, or of his limited interactional experience, or of the natural distributional properties of linguistic patterns, would not be consistent with experimental outcomes suggesting that there is no (or little) evidence of schematic (full) productivity in the elicited production of his peers. Hence, there would be no evidence suggesting abstract representation; spontaneous production is accountable for in lexically-specific terms and the child's peers in the experimental study do not show adult-like productivity with nonce verbs (which would be evidence of more schematic representation). If the child in the naturalistic study is using abstract competence to produce (or assemble) sentences that can be (mostly) accounted for in lexically-specific terms, why wouldn't his peers provide evidence of such schematic competence when prompted in experimental settings?

Similarly, claiming that lack of productivity in experimental settings is caused by ancillary factors (e.g. task difficulties) would be problematic in light of naturalistic results suggesting that most of what the child says can be accounted for in lexically-specific terms. If, for instance, two-year-olds do not show adult-like productivity because of ancillary factors, why isn't there evidence of such more schematic representation in the spontaneous production of one of their peers? Thus, converging evidence would suggest that two-year-olds are not as productive as adults in experimental settings because at that age children still rely on lexically-specific patterns acquired from their concrete language input.

Thus, the design of this study aims at providing a picture as exhaustive as possible on the nature of early Italian and whether it can be accounted for in lexically-specific terms.

Part II

The Spontaneous Production of a two-year-old child



Page intentionally left blank

5.

INTRODUCTION

The following longitudinal study collects data on the spontaneous production of an Italian-speaking child and aims to understand whether and how much of Italian-speaking children's language can be accounted for by lexically-specific units. In order to do so, it adopts the traceback method (Lieven et al., 2003, 2009; Dąbrowska & Lieven, 2005), whose rationale has been discussed in the previous chapters.

The specific research question that is investigated is the following:

> *Can Italian-speaking children's early language be accounted for in terms of lexically-specific units acquired from the concrete language that children themselves have previously experienced?*

Page intentionally left blank

6.

METHOD

6.1. PARTICIPANTS

One male child, whose pseudonym is Roberto, was followed for six consecutive weeks from 2(years);1(month).13(days) to 2;2.26 of age.

The child lived in a small village in Trentino (North-East of Italy). His parents were middle-class, university-educated professionals, whose families belonged to working class backgrounds (farmers and factory labourers). Both parents, whose pseudonyms are Sara (mother) and Sebastiano (father), were amateur musicians and often encouraged Roberto to sing. Roberto was an only child.

Other people present during the recordings were grandparents, the researcher himself, Roberto's cousin Federico (one year younger than him), uncles and aunts, family friends and an unidentified female speaker.

The family was recruited through mutual acquaintances. Once the family showed interest in participating, the sampling regime to be used and the methodology of the research was explained. Consequently, they knew that the research was about language acquisition and had a vague idea of the traceback method to be adopted.

Throughout the length of the study, paternal grandparents overwhelmingly used the local dialect (called *Trentin*) when they spoke to each other and to Sebastiano (who in turn used Trentin with them), whereas they used Italian when addressing Roberto. Instances of code switching depending on to whom grandparents were talking are frequent in the recordings in which they appear (see Appendix_II, ch. 20.1, for an example).

Sara (mother) and Sebastiano (father) addressed their son only in Italian throughout the whole study, with the exception of a few Trentin words, such as *vara* (Italian *guarda*: look-IMP.2.SG). In a few, extremely rare occasions, Sebastiano addressed his son with Trentin. There is no clear sign of the use of Trentin words or structures in Roberto's speech (except for the word *vara*).

6.2. DATA COLLECTION

6.2.1. RECORDINGS

Recordings took place at Roberto's home, at his paternal grandparents' home, and on one occasion, at an outdoor playground. Each recording session lasted between fifteen minutes and two hours and seven minutes. The child was recorded for approximately five hours a week for six weeks in a row and the total corpus amounts to thirty-six hours and eight minutes. The first recording took place when Roberto was 2;1.15 and the last one when he was 2;2.26.

When the researcher was present (once a week) he conducted the recordings, when he was not, the family did. Sessions were recorded with an Olympus LS-5 voice recorder.

The activities in which Roberto was involved were playing and chatting with the researcher, his parents, grandparents, aunts and cousins, reading books and telling stories, singing, helping parents and grandparents to cook and having snacks and meals with the extended family.

6.2.2. VOCABULARY QUESTIONAIRRE

Towards the end of week 5 (when Roberto was twenty-six months and eighteen days old) the parents completed the Italian version of the MacArthur Form Vocabulary Check List: Words and Sentences (Caselli & Casadio, 1995). In accordance with Caselli and Casadio's (1995) guidelines, the parents were given the questionnaire a couple of weeks before filling it in. They were advised to read through the questions and observe the child for a few days before completing the questionnaire. They were also advised not to take more than one week to complete it.

The results of the questionnaire showed that Roberto was a very advanced learner. His parents reported that he knew 582 words, nearly one and a half times as many words as the average vocabulary size at age twenty-six/seven months (334 words; sd 174). Roberto's scores were well above the 90th percentile of his age group (about 550 words). Indeed, his score was well above the average vocabulary size reported for children as old as thirty months (446 words). In fact, it was above the 75th percentile of 580 words reported by Caselli and Casadio (1995) for thirty-month-olds.

Roberto was also advanced when it came to measures of grammatical complexity. His scores showed that his linguistic development was more in line with the one displayed by thirty-month-old children than with that of his average peers. Tab 6.1 summarises the various measures of grammatical complexity tested by the MacArthur questionnaire. Roberto's results are compared with those of both his peers and thirty-month-olds, as reported by Caselli and Casadio (1995).

Table 6.1: more qualitative measures of the MacArthur Questionnaire. Roberto's results are compared with both his peers' and thirty-month-olds' average results.

LANGUAGE WITHOUT EXTRA-LINGUISTIC CONTEXT	Roberto	26-27-month-olds who often do so	30-month-olds who often do so
is the child able to understand language without extra-linguistic context? (about objects and people not present, future and past events)	often	70%	90%
is the child able to speak about things without extra-linguistic context? (about objects and people not present, future and past events)	often	50%	more than 70%
MORPHOLOGICAL COMPLEXITY	**Roberto**	**proportion of 26-27-month-olds who show morphological complexity**	**proportion of 30-month-olds who show morphological complexity**
does the child use both singular and plural forms of the same noun?	yes	70%	87.5%
does the child use various forms of the same adjective (number and gender inflections)?	yes	55%	81.3%
does the child use various forms of the verb (mood-tense-person markers)?	yes	67.5%	87.3%
GRAMMATICAL COMPLEXITY	**Roberto**	**26-27-month-olds**	**30-month-olds**
use of functional words, adverbial phrases, co-ordination and subordination (*daddy went away with the car)* as opposed to paratactic strategies (daddy away car) - % of complex sentences	100%	less than 50%	less than 80%
use of pronouns (*I want this*) as opposed to nouns (*Roberto wants this*) - % of pronoun usages	66%	more than 30%	more than 50%

6.2.3. TRANSCRIPTIONS

All recordings were transcribed using standard Italian traditional spelling and following CHAT conventions in accordance with MacWhinney's (2000) guidelines. The researcher transcribed all recordings and double checked two minutes of each recording. Transcriptions were accurate, with the exception of a few typographical errors that were amended. Each hour of recording took twenty to fifty hours of work to be transcribed. First names and surnames were modified with pseudonyms. The only exception was the researcher's name, whose real name (Luca) was not modified.

6.3. PROCEDURE

6.3.1. TEST AND MAIN CORPUS

The corpus collected was divided into two corpora: **test** corpus and **main** corpus. The **test corpus** consists of the transcriptions of the last two recordings, which were taken on the very last day of the study. The first file amounts to one hour and five minutes and was recorded during lunch time by the mother. The second file amounts to two hours and two minutes and was recorded by both the mother and the researcher between 4 pm and 6 pm approximately. The whole test corpus amounts to three hours and seven minutes. The **main corpus** amounts to thirty-three hours and one minute and consists of all remaining recordings.

6.3.2. IDENTIFYING INTELLIGIBLE MULTI-WORD SENTENCES

Each multi-word sentence the child produced in the **test corpus** was singled out. In total there were 1890 sentences, of which 1489 (79%) contained two or more words.

Sentences that contained parts that were not intelligible - transcribed as xxx, (1b) - or contained an uncertain transcription of a word - transcribed as <word> [?], (1a) - were discarded.

1. a) **CHI:* *<l' ho trovato> [?].*

l'	*ho*	*trov-a-t-o*
clitic.3.ACC	have(PRS.1.SG)	find-TV(conj.I)-PTCP-M.SG

"I found it."

(week6.2014.02.18.B: line 393)

b) **CHI:* *<non xxx> [=! whispering].*

non xxx

not xxx

"Don't xxx."

(week6.2014.02.18.B: line 2729)

A further 106 sentences were excluded from the analysis because the child's sentences had been interrupted, either by another speaker or by the child himself. These sentences were discarded because it was not possible to establish what exactly Roberto wanted to say. In (2) the sentence could potentially be continued by, for example, an adjective (*little boots are clean*) or a location (*little boots are there*). Thus, (2) was excluded because it was not possible to establish its putative precedent units (not in a satisfactory way, at least). In (2) it is not clear whether Roberto's sentence was to be traced back to strings instantiating, for instance, the schema *THINGS are QUALITY* (i.e. he wanted to say *little boots are clean*) or the schema *THINGS are PLACE* (i.e., he wanted to say *little boots are there*).

2. **CHI: stivaletti so(no) +//.*
   Stival-ett-i so(no) +//.
   Boot-little(DIM)-M.PL be(PRS.3.PL)
   "Little boots are +//."
   (week.6.2014.02.18.B: line 1224)

Whenever possible however, the child's sentence was considered for analysis. Sentence (3) is a co-ordination of two clauses. The second clause is self-interrupted and hence discarded. Nevertheless, the first clause is intelligible and is considered for analysis. Hence, what is analysed is not the whole sentence but, instead, the underlined part below (namely the first clause).

3. **CHI: voio [: voglio] [* p] &*RES:sì attaccare (.) <un> [//] questo e voio [: voglio] [* p] +//.*
   voglio attaccare questo e voglio +//.
   vogli-o attacc-a-re quest-o
   want-PRS.1.SG attach-TV(conj.I)-INF this-M.SG
   e vogli-o +//.
   and want-PRS.1SG
   "I want to attach this and I want +//."
   (week6.2014.02.18.B: lines 1334-1335)

This procedure yielded 993 multi-word sentences that were intelligible. Table 6.2 shows that 52% of Roberto's production consisted of intelligible multi-word sentences.

Table 6.2: Roberto's test corpus production: identifying multi-word sentences.

sentence type	no.	%
intelligible multi-word sentences	993	52%
one-word sentences	401	21%
non intelligible multi-word sentences	390	21%
interrupted sentences	106	6%
total	**1890**	**100%**

6.3.3. IDENTIYFYING NOVEL SENTENCES TO BE ANALYSED (TARGET SENTENCES)

A further step was taken in order to tell apart those sentences that could be considered novel (and hence creative) from those sentences that could be thought of as non-creative.

Songs, poems and nursery rhymes (seventy sentences) were discarded from the analysis, as their formulaic and mnemonic nature hardly allowed for them to be thought of as creative usages of language.

Similarly, imitations of what adults had just uttered and repetitions of what Roberto himself had just said were not considered because of their "parroting" nature. Two kinds of imitations and self-repeats were identified: partial and full.

Full Imitations and **full self-repeats** are those sentences the child produced that share the exact same words and morphemes, independent of their linear order, with other speakers' (full imitation, 4) or the child's (full self-repeat, 5) sentences found in any of the five previous lines of the transcription. All full-imitations (forty-two sentences) and the only full self-repeat presented the same linear order as their immediate precedents.

4. a) *MOT: *un'altra?*
un' altr-a ?
a other-F.SG ?
"Another one?"

b) *CHI: *un'altra*
un' altr-a
a other-F.SG
"Another one."

(week.6.2014.02.18.A.: lines 84 and 85).

5. a) *CHI: *col motorino.*
co=l motorin-o
with=the(M.SG) scooter-M.SG
"With the scooter."

b) *CHI: *col motorino.*
co=l motorin-o
with=the(M.SG) scooter-M.SG
"With the scooter."

(week6. 204.02.18.B: lines 76 and 81)

Reduced Imitations and **reduced self-repeats** are those sentences the child uttered that correspond to a **continuous** string of words which constitutes a subpart of a sentence any other speaker (reduced imitation, 6b) or the child himself (reduced self-repeat, 7b) uttered in one of the previous five lines. The whole target sentence produced by the child and the subpart of the preceding sentence must share the exact same words and morphemes, independent of the linear order of their constituents. None of reduced imitations (ten sentences) or self-repeats (twenty-four sentences) presented linear orders other than the ones attested in their precedents.

6. a) *RES: *no*, *questa* ***è una salciccia.***
no, quest-a ***è un-a salsicci-a***
no, this-F.SG **is a-F.SG sausage-F.SG**
"No, this **is a sausage**."

b) *CHI: ***è una salciccia.***
è un-a salsicci-a
is a-F.SG sausage-F.SG
"(It) **is a sausage**."

(week6.2014.02.18.B: lines 1692 and 1693)

7. a) *CHI: ***lo porta via*** *il chiamoncino [: camioncino] [* p].*
l-o ***port-a*** ***via*** *il*
clitic.3.ACC-M.SG **take-PRS.3.SG** **away** the(M.SG)
camion-cin-o
truck-little(DIM)-M.SG
"The little truck, (he) **takes it away**."

b) *CHI: ***lo porta via.***
l-o ***port-a*** ***via***
clitic.3.ACC-M.SG **take-PRS.3.SG** **away**
"(He) **takes it away**."

(week6.2014.18.A: lines 1369 and 1370)

Note, that in order for sentence B to be considered as an imitation or self-repeat of sentence A, the former **must** either present the exact same words and morphemes as the **whole** sentence A (full imitation and full self-repeat) or correspond to a **continuous** subpart of it (partial imitation and self-repeat), with which the whole sentence B shares the exact same words and morphemes.

In (8b) below, the child expands the adult's utterance (8a); hence it is not an imitation.

8. a) *RES: *ah@i, ha scavato e ha tirato sù un tubo.*
ha *scav-a-to* *e* *ha* *tir-a-to*
has dig-TV(conj.I)-PTCP and has pick-TV(conj.I)-PTCP
sù *un* *tub-o*
up a(M.SG) pipe-M.SG
"(He) dug and picked (pulled) up a pipe."

b) *CHI: *ha scavato e ha tirato (.) sù un tubo e ha fatto tcsh@o.*
ha *scav-a-to* *e* *ha* *tir-a-to*
has dig-TV(conj.I)-PTCP and has pick-TV(conj.I)-PTCP
sù *un* *tub-o* *e* *ha* *fatto* *tcsh*
up a(M.SG) piper-M.SG and has done *tcsh*
"(He) dug and picked (pulled) up a piper and went 'tcsh'."

(week6.2014.02.18.B: lines 3098 and 3099)

In (9b) overleaf, Roberto's sentence presents the same number of words and morphemes as its precedent (9a). However, he changes the verb form of the adult's sentence: *ho* "have (PRS.1.SG)" becomes *hai* "have (PRS.2.SG)". (9a) and (9b) do **not** share the exact **same** words and morphemes and hence the latter is not an imitation of the former.

9. a) *RES: *<l' ho> [/] l' ho visto uno morto.*
l' *ho* *vis-t-o*
clitic.3.ACC have(PRS.1.SG) see-PTCP-M.SG
un-o *mort-o*
one-M.SG dead-M.SG
"I've seen a dead one."

b) *CHI: *l' hai visto uno morto.*
l' *hai* *vis-t-o*
clitic.3.ACC have(PRS.2.SG) see-PTCP-M.SG
un-o *mort-o*
one-M.SG dead-M.SG
"You've seen a dead one."

(week6.2014.02.18.B: lines 2616 and 2617)

Both (8b) and (9b) are therefore considered for analysis.

Importantly, if the child's utterance does not correspond to a **continuous** string of words in the preceding sentence, this is not considered an imitation or self-repeat (be it partial or full). In the fictitious (10b), the child does **not** reuse the ***whole*** **continuous string**, as he drops *ieri* "yesterday". As a result, the **number** of words does not match and (10b) is **not** considered an imitation of (10a).

10. a) *ADL: *il papà ieri è andato a lavorare*
il *papà* *ieri* *è* *and-a-t-o*
the(M.SG) daddy(M.) yesterday is go-TV(conj.I)-PTCP-M.SG
a *lavor-a-re*
to work- TV(conj.I)-INF
"Yesterday Daddy went to work."

b) *CHI: *il papà è andato a lavorare*
il *papà* *è* *and-a-t-o* *a*
the(M.SG) daddy(M.) is go-TV(conj.I)-PTCP-M.SG to
lavor-a-re
work- TV(conj.I)-INF
"Daddy went to work."

Finally, when the child's utterance is an imitation/self-repeat (be it partial or full) of a sentence that is located more than five lines before it, the child's sentence is considered new and is considered for analysis (11b).

11. a) *MOT: *e poi prende l' aereoplano e* ***va <lontano>*** *[/] lontano „ vero?*

e	*poi*	*prend-e*	*l'*	*areoplan-o*	*e*
and	then	take-PRS.3.SG	the	airplane-M.SG	and
va	***lontano***	*lontano,*	*vero?*		
goes	**far**	far ,	true?		

"And then (he) takes the airplane and goes far away, right?"

b) *CHI: **va lontano**

va	***lontano.***
goes	**far.**

"(He) goes away."

(week6.2015.02.18.A: lines 542 and 551)

A final procedure was used to identify the **sentence types** uttered by Roberto. Test corpus sentences he uttered in the exact same way on more than one occasion were considered different instances of the same sentence type and only the first occurrence was considered. Thus, if the child uttered two sentences which shared the exact same words and morphemes, these were considered instances of the same sentence type. As Italian has a fairly free WO, this was also the case when the different sentences varied in the way constituents were ordered (the way flexible word order (WO) has been handled is discussed in detail in 6.3.4 and 6.4.1.1). Hence, both (12a) and (12b) are occurrences of the same sentence type and only the first one is analysed.

12. a) *CHI: *il lupo mangia tutto.*

il	*lup-o*	*mangi-a*	*tutt-o*
the(M.SG)	wolf-M.SG	eat-PRS.3.SG	all-M.SG

b) *CHI: *mangia <tutto> [!] il lupo.*

mangi-a	*tutt-o*	*il*	*lup-o*
eat-PRS.3.SG	all-MSG	the(M.SG)	wolf-M.SG

"The wolf eats everything."

(week6.2014.02.18.B: lines 2825 and 2831)

Seventy-three sentences matched an already attested target sentence in terms of identity of words and morphemes **and** the linear order in which these appeared. A further five sentences were identical to previously analysed sentences in terms of the words and morphemes that appeared in them, but presented a different linear order (as in 12).

This procedure yielded 768 sentence types, which represent the **target sentences** that had been analysed and that represent the focus of this longitudinal study. Table 6.3 shows that 77% of Roberto's intelligible multi-word sentences could be considered novel.

Table 6.3: Roberto's intelligible multi-word-sentences: identifying target sentences.

Roberto's intelligible multi-word sentences	no.	%
novel multi-word sentences (target sentences)	768	77%
instantiations of an already considered target sentence type	78	8%
songs, poems and nursery rhymes	70	7%
partial and full imitations	52	5%
partial and full self-repeats	25	3%
TOTAL	**993**	**100%**

6.3.4. IDENTIFYING COMPONENT UNITS

Each target sentence was then traced back by identifying the closest string(s) of words in the **main corpus** that matched it. Matching strings are putative **component units** of a particular target sentence (refer back to 3.5.2).

Component units are morphemes, words or **continuous** strings of words **attested in the main corpus** which share morphological and/or lexical material with the target sentence.

There are two types of component units: **Fully Lexically-specific Strings** and **Schemas-with-Slots**. Component units **need not** occur in isolation but **must**

> "correspond to a chunk of semantic structure"
>
> (Dąbrowska & Lieven, 2005, p. 47),

such as *PROCESS*, *THING*, *QUALITY, PLACE, GIVER, RECEIVER* and so forth.

6.3.4.1. Fully Lexically-specific Strings

A **Fully Lexically-specific String** is a word or **continuous** string of words attested in the **main corpus** which shares the exact same words and morphemes, **independent of their linear order**, with the target sentence (or part of it).

In order to be considered as an available component unit, a (putative) *Fully Lexically-specific String* has to appear in the main corpus at least twice, excluding imitations and self-repeats (refer back to 6.3.3 for what is classified as imitation and self-repeat) and can be attested in any speaker's speech.

Fully Lexically-specific Strings can be thought of as pre-packed strings of various length. Because Italian is a language which presents a flexible word order (WO), multi-word *Fully Lexically-specific Strings* can have either a flexible or a fixed internal order of constituents. Loosely following Dąbrowska (2014), I refer to the former (flexible WO) as **fully-specific-packets** and to the latter (fixed WO) as **fixed-strings**.

In order for a **fully-specific-packet** to be considered as an available component unit, the same string (same **continuous** "block" of words and morphemes corresponding to a coherent chunk of semantics) must be attested at least twice in the main corpus with at least two different internal orders of constituents.

The parts highlighted in **bold** of (14b-c) are instantiations of the fully-specific-packet (14a)[39], to which the emboldened part of target sentence SB427 (13) was traced back. The two instantiations of (14a), i.e. (14b-c), and the part of the target sentence highlighted in **bold** share the exact same words and morphemes. Since the same **continuous** "block" of words and morphemes corresponds to a unified and coherent chunk of semantics (*PROCESS*) and also presents different linear orders, the method assumes that (13), (14b) and (14c) are all instantiations of (14a).

13. TARGET SENTENCE
*CHI: *no, qui,* ***così non si fa!***
no, qui, ***così non si fa***
no, here **like.that not clitic.IMPRS.NOM does**
"No, here, that's not the way to do it."
(week6.2014.02.18.B: line 3088)

[39] Refer back to 1.3.5 for the deontic use of *si* impersonal subject used in these examples.

14. UNIT AND ITS INSTANTIATIONS

a) ***Così*** + ***non<si<fa***

così ***non*** ***si*** ***fa***
like.that **not** **clitic.IMPRS.NOM** **does**
"That's not the way to do it."

b) *FAT: *digli (..)* ***così non si fa .***
d-i=gli ***così*** ***non***
tell-IMP.2.SG=clitic.3.DAT **like.that** **not**
si ***fa***
clitic.IMPRS.NOM **does**
"Tell him, that's not the way to do it."
(week1.2014.01.12.B.chat: line 180)

c) *CHI: ***non si fa così !***
non ***si*** ***fa*** ***così***
not **clitic.IMPRS.NOM** **does** **like.that**
"That's not the way to do it."
(week3.2014.01.21: line 2081)

However, WO variability within *Fully Lexically-specific Strings* is not assumed and must be found amongst the instantiations of the particular unit at stake. The unit used in 13 (namely 14a) is attested in the main corpus with two different orders (14b and 14c) and therefore it is assumed that such variability is available to the child. If a particular string appears only with a specific linear order, it is assumed that the unit does not present WO variability (hence it is a **fixed-string**).

For example, target sentence SB313 (15) presents the string *lì va* "there goes", which appears in the main corpus only as *va lì* "goes there"; (16a and 16b).

Consequently, the highlighted part of the target sentence in (15) cannot be traced back to the strings in (16) (as the only order available in the test corpus is *va lì* "goes there"), even though all strings share the exact same words and morphemes.

15. TARGET SENTENCE

*CHI: *e* ***lì va*** *que [: questo] [* p].*
e ***lì*** ***va*** *quest-o*
and **there** **goes** this-M.SG
"And this one goes there."
(week6.2014.02.18.B: line 2363)

16. PUTATIVE UNITS

a) *FAT: *<va lì> [<]* !
va lì
goes there
"(It) goes there."
(week6.2014.02.12: line 1197)

b) *MOT: *eh@i , e va lì .*
eh, e va lì
eh, and goes there
"Eh, and (it) goes there."
(week3.2014.01.21: line 2698)

Such a distinction between **fixed-strings** and **fully-specific-packets** is codified by using a plus sign (+) whenever the internal order is not fixed. Therefore, the unit that appears in (13) and (14) is reported as (17a) (flexible WO). The sign plus (+) indicates that there is WO flexibility. The symbol "<" indicates that certain elements in the string have fixed order. Hence, the interaction of "<" and "+" indicates that in (17a) *così* "like that" may either precede or follow the string *non si fa* (indeed the same unit can be codified as 14a or 17a). The fixed-string instantiated in (16a-b) is instead represented as in (17b); i.e. when fixed-strings are reported, their elements are just reported with the order in which they appear (17b).

17. a) FULLY-SPECIFIC PACKET

non<si<fa + così
non si fa così
not clitic.IMPRS.NOM does like_that
"That's not the way to do it."

b) FIXED STRING
va lì
va lì
goes there
"(It) goes there."

When a *Fully Lexically-specific String* is a sequence of words, no other element can appear within it. Target sentence SB394 is reported in (18). The highlighted string of target sentence SB394 (18) (*il Luca va a casa* "Luca goes home") appears twice in the main corpus (19a and 19b).

18. *CHI: *tra poco <il lupo [: Luca] [* s:ur] > [//]* ***il Luca vas [: va] [* p] <ca [: casa] [* p -ret]> [//] a casa***
Tra poco ***il Luca va a casa***

tra	*poco*	***il***	***Luca***	***va***	***a***	***cas-a***
between	little	**the(M.SG)**	**Luca**	**goes**	**to**	**home-F.SG**

("Luca goes home in a little while").
"Luca will go home in a while/ Luca is going home soon."
(week6.2014.18.B: line 2889-2890)

19. a) *MOT: *il Luca va a casa !*

il	***Luca***	***va***	***a***	***cas-a***
the(M.SG)	**Luca**	**goes**	**to**	**home-F.SG**

"Luca goes home."
(week5.2014.02.04: line 2006)

b) *MOT: ***il Luca*** *adesso* ***va a casa***

il	***Luca***	*adesso*	***va***	***a***	***cas-a***
the(M.SG)	**Luca**	now	**goes**	**to**	**home-F.SG**

"Now, Luca is going home."
(week2.2014.01.14.A: line 193)

In (19b), the fixed-string is not continuous as *adesso* "now" is inserted between *Luca* and *va* "goes". Hence, (19a) is an instance of the fixed-string *Luca va a casa* "Luca goes home", but (19b) is not.

To summarise, a *Fully Lexically-specific String* is a word or a **continuous** string of words attested in the **main corpus** which shares the exact same words and morphemes with the target sentence (or part of it). A specific *Fully Lexically-specific String* can have either a fixed (**fixed-strings**) or a flexible (**fully-specific-packets**) order of constituents. In the former case it is codified as *specific string*, in the latter as *specific + string*. In order to be considered as an available component unit, a *Fully Lexically-specific String* must be found in the main corpus at least twice, **excluding imitations and self-repeats** (see 6.3.3.) and can be attested in either the child's (14c) or any other speaker's (14b) main corpus strings.

6.3.4.2. Schemas with Slots

A **Schema-with-Slot** is a component unit which can be instantiated by any word/string of words that shares some recurring lexical and/or morphological material with the target sentence, but with variation on the same morphological ending(s) and/or word(s)/string(s) of words. The invariable elements shared by the schema and its instantiations must present the exact same words and morphemes (fig. 6.1 and 6.2, highlighted in blue). The variable elements that appear in such otherwise identical strings are called **slots** (fig. 6.1-6.2, highlighted in white) and each schema can contain up to two slots. Slots represent semantic (fig. 6.1) and/or morphological (fig. 6.2) generalisations inferred by drawing analogies between the variable concrete material that appears in the different concrete strings (fig. 6.1 and 6.2, green strip) which instantiate the schema (fig. 6.1-6.2, yellow strip). It follows that slots and their concrete fillers must share certain properties; they are (fillers) or represent (slots) either morphological inflections (be they verbs' aspect-mood-tense-person-number markers or nouns' inflectional endings) or words/strings of words that correspond to the same broad semantic role (agent, patient, action, place etc.) in the schema. Schemas-with-Slots must

> "correspond to a chunk of semantic structure"
>
> (Dąbrowska & Lieven, 2005, p. 447)

and must be attested at least twice in the main corpus, excluding imitations and self-repeats. In order to create a slot, two or more different fillers that share the same semantic and/or morphological features must occur within the same otherwise identical expression (see white parts on green strips, fig. 6.1-6.2).

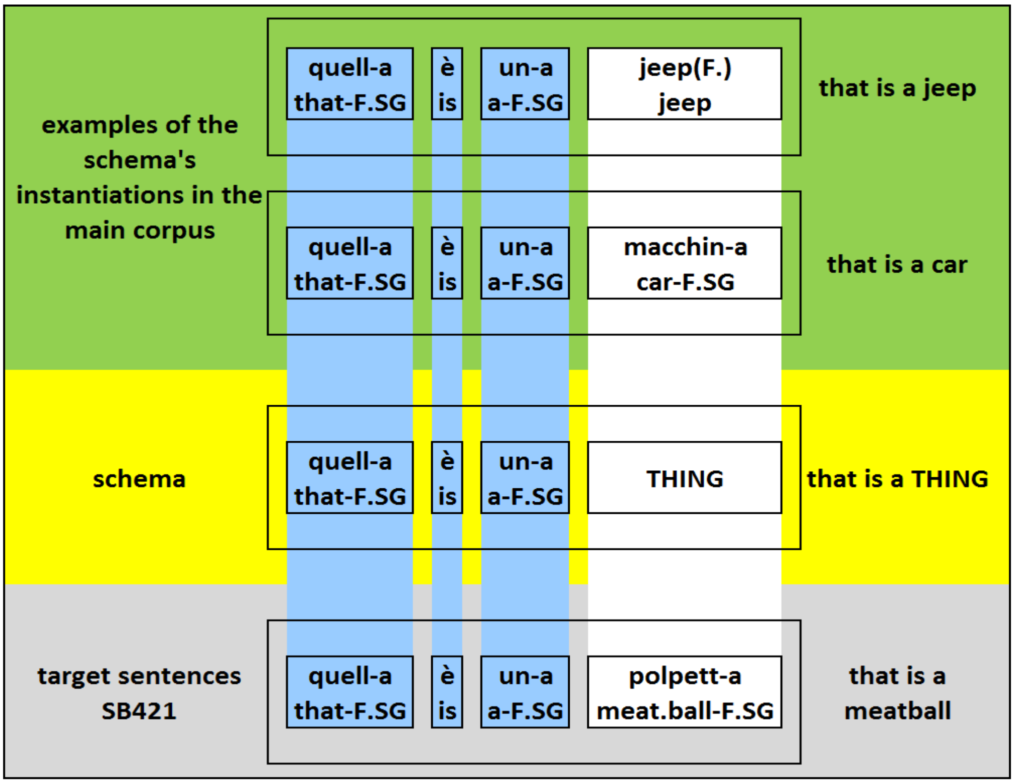


Figure 6.1: target sentence (grey strip), schema (yellow strip) and the schema's instantiations (green strips). The slots and its instantiations are highlighted in white, whereas the lexically specific part of the schema is highlighted in blue.

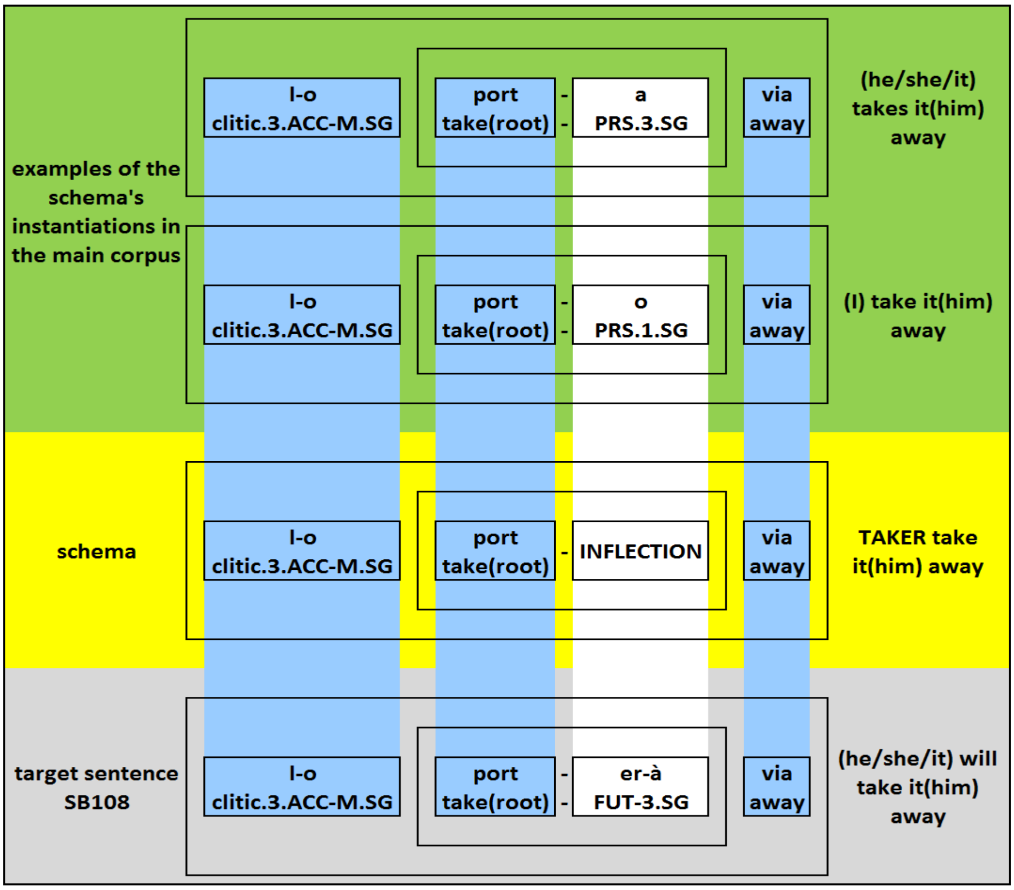


Figure 6.2: see fig. 6.1 above on how to read this figure.

Schemas and their instantiations **need not** share the same order of constituents. A Schema-with-Slot which is attested in the main corpus with at least two different internal linear orders is considered to be a component unit whose internal WO is flexible. A Schema-with-Slot which is attested in the main corpus with one linear order only is considered to be a component unit whose internal WO is fixed. As with *Fully Lexically-specific Strings*, when a schema is flexible with respect to the way its constituents are ordered, it is codified with a sign plus (+). The schemas in fig. 6.1-6.2 (yellow strip) have a fixed word order and are reported as in (20a-b). Conversely, (21a), which is reported with “+” and “<”, has a flexible order of constituents, as its instantiations (21b-c) present WO variability. I refer to (20a-b) as **fixed-schemas** and to units like (21a) as **schematic-packets**.

20. FIXED-SCHEMAS
a) *Questa è una THING*
quest-a   è   un-a   THING
this-F.SG   is   a-F.SG   THING
“This is a THING.”

b) *Lo port-INFLECTION via*
l-o   port-INFLECTION   away
clitic.3.ACC-M.SG   take(root)-INFLECTION   away
“TAKER take it away.”

21. SCHEMATIC-PACKETS
a) La<THING + è<caduta [S+V]
l-a   THING   è   cad-u-t-a
the-F.SG   THING   is   fall-TV(conj.II)-PTPC-F.SG
“The<THING + has<fallen.”

*b)*CHI: è caduta la tenda* [VS]
è   cad-u-t-a   l-a   tend-a
is   fall-TV(conj.II)-PTPC-F.SG   the-F.SG   curtain-F.SG
“The curtain has fallen.”
(week2.2014.01.16: line 889)

*c)*RES: la torre è caduta* [SV]
l-a   torr-e   è   cad-u-t-a
the-F.SG   tower(F.)-SG   is   fall-TV(conj.II)-PTPC-F.SG
“The tower has fallen.”
(week3.2014.01.21: line 1889)

As with *Fully Lexically-specific Strings*, a schema must be attested at least twice in the main corpus, excluding imitations and self-repeats (partial or full). This means that both imitations and self-repeats (full or partial) are discarded and not considered as instantiations of either type of unit.

Full and partial self-repeats are those words/continuous strings of words which share the exact same words and morphemes, independent of their linear order, with a sentence (full self-repeat) or part of a sentence (partial self-repeats) uttered by the *same* speaker in one of the previous five lines of the transcription. When the previous sentence is uttered by *another* speaker, it is a case of imitation (full or partial).

So, for example, if the Fully Lexically-specific String searched were (22), the father's sentences in (23a-b) could be classified as instantiations of it. Both sentences (23a-b) share the exact same *continuous* "*block*" of words and morphemes which corresponds to a coherent chunk of semantics (namely *PROCESS*). Consequently, they can be thought of as instantiations of the same unit, namely (22).

22. a) *Il<treno + sta<per<partire*
il *tren-o* *st-a* *per*
the(M.SG) train-M.SG stay-PRS.3.SG to
part-i-re
leave-TV(conj.III)-INF
"The train is about to leave."

23. a) *FAT: *il treno sta per partire*
il *tren-o* *st-a* *per*
the(M.SG) train-M.SG stay-PRS.3.SG to
part-i-re
leave-TV(conj.III)-INF

b) *FAT: *sta per partire il treno*
st-a *per* *part-i-re* *il*
stay-PRS.3.SG to leave-TV(conj.III)-INF the(M.SG)
tren-o
train-M.SG

"The train is about to leave."

(week5.2014.02.07: lines 2541 and 2542)

However, the two sentences are uttered one after the other. As a result, (23b) is discarded as **self-repeat** of (23a). Only the latter is considered as an instantiation of (22). Had (23b) been uttered by a speaker other than the father, it would have been a case of **imitation** rather than self-repeat. Either way, the second sentence would be discarded and not considered in the counting. It follows that, in order for (22) to be considered as an available unit, the string must be instantiated at least one other time in the main corpus. If any other such instance has the same order as (23a), then the unit is a **fixed-string** (*il treno sta per partire*; "the train is about to leave"), as both strings present the same linear order. If the second instance has a different WO (as in 23b), the unit attested is a **fully-schematic-packet** (as coded in 22).

A Schema-with-Slot has some elements (*SLOTS*) that are not lexically specified, which implies that two different instantiations of the same schema do not necessarily share the exact same words and morphemes (as the slot may be instantiated by different fillers). What is classified as a self-repeat or imitation is a main corpus instantiation (*that's a car)* of a putative schema (*that's a THING*) which is identical to another instantiation (*that's a car*) of the same schema uttered in the previous five lines. That is, self-repeats and imitations are main corpus sentences which share the exact same word and morphemes, **independent of their linear order**, with another instantiation of a putative schema appearing in one of the previous five lines.

For instance, the fictitious (24) could be the target sentence to be traced back. The target sentence (24) would be traced back to (25b), which is the closest main corpus string that matches it. Both (24) and (25b) are instantiations of the putative schema in (26), which is also instantiated by (25a).

However, (25b) is a full self-repeat of (25a), since the two share the exact same words and morphemes (even though they have different WOs). Consequently, only (25a) would be considered as an instantiation of (26), to which both the target sentence (24) and the units in the main corpus (25) could be traced back.

24. *CHI: *voglio questo*
    vogli-o *quest-o*
    want-PRS.1.SG this-M.SG
    "I want this one."
    (fictitious example)

25. a) *CHI: *il pane voglio*
il *pan-e* *vogli-o*
the(M.SG) bread(M.)-SG want-PRS.1.SG
"I want bread."
(fictitious, line 1)

b) *CHI: *voglio il pane*
vogli-o *il* *pan-e*
want-PRS.1.SG the(M.SG) bread(M.)-SG
"I want bread."
(fictitious, line 5)

26. *Voglio + THING_WANTED*
vogli-o + *THING_WANTED*
want-PRS.1.SG + THING_WANTED
"I want THING_WANTED."

A slot is created when the lexically specific part of the schema (*voglio* "I want") is found in combination with at least two different fillers that correspond to the same broad semantic role (THING_WANTED). Consequently, another instantiation of the schema in which the SLOT is elaborated by an item other than *il pane* "the bread" is needed for the schema to be considered as an available unit. From this standpoint, it makes no difference whether (25b) is an available instantiation of the schema or not.

Nevertheless, if a further instantiation of the schema were (27), both instantiations (25a and 27) would present the linear order *THING_WANTED–I-want*. It follows that the schema attested in the main corpus would not be the schematic-packet hypothesised in (26) (*I-want* + *THING_WANTED*), but the fixed-schema in (28), whose constituents have a fixed WO.

27. **CHI: quella voglio*
quell-a vogli-o
that-F.SG want-PRS.1.SG
"I want that one."
(fictitious example)

28. *THING_WANTED voglio*
THING_WANTED *vogli-o*
THING_WANTED want-PRS.1.SG
"I want THING_WANTED."

No word order variation is found in the main corpus, and so it is assumed that the schema available to the child does *not* have a flexible internal order. The target sentence (24) would not be able to be traced back to (25a) and (27), as it presents the different order *I-want-THING_WANTED*. If (25b) were **not** an imitation of (25a) and could be considered as an instantiation of (26), the schema in (26) would be considered an available unit. This would allow (24) to be traced back to (26) - whose instantiations would then be (25a-b) and (27).

To summarise, Schemas-with-Slots are words/strings of words that share some recurring lexical and/or morphological material with a target sentence, but with variation on the same morphological ending(s) and/or word(s)/string(s) of words. Variable elements are lexically unspecified elements called SLOTs. The latter represent generalisations from the variable concrete material that appears in the different strings which instantiate the schema. A slot is created when the lexically specific part of the schema is combined with at least two different fillers that instantiate that particular slot. Schemas-with-Slots can have either a fixed (fixed-schemas) or a flexible (schematic-packets) linear order of constituents and must be attested at least twice in the main corpus, excluding imitations and self-repeats.

6.3.5. DERIVING TARGET SENTENCES: TYPES OF OPERATIONS

Once its putative component units have been identified, a target sentence can be derived by applying two types of operations to these (available) component units: juxtaposition and superimposition.

Juxtaposition is the operation though which a component unit_A (29c) is added to either end of another component unit_B (29b) in order to derive the target sentence (29a), as illustrated in fig. 6.3.

29. a) *CHI: ***stava <lavorando (..)> [>] lassù.***

st-a-v-a | *lavor-a-ndo*
stay-TV(conj.I)-IMPERF-3.SG | work-TV(conj.I)-ing
la+(s)sù
there+up
"(He/she/it) was working up there."
(week6.2014.02.18.A: line 1192)

b) *Stava lavorando*
st-a-v-a | *lavor-a-ndo*
stay-TV(conj.I)-IMPERF-3.SG | work-TV(conj.I)-ing
"(He/she/it) was working."

c) ***Lassù.***
la+(s)sù
there+up
"Up there."

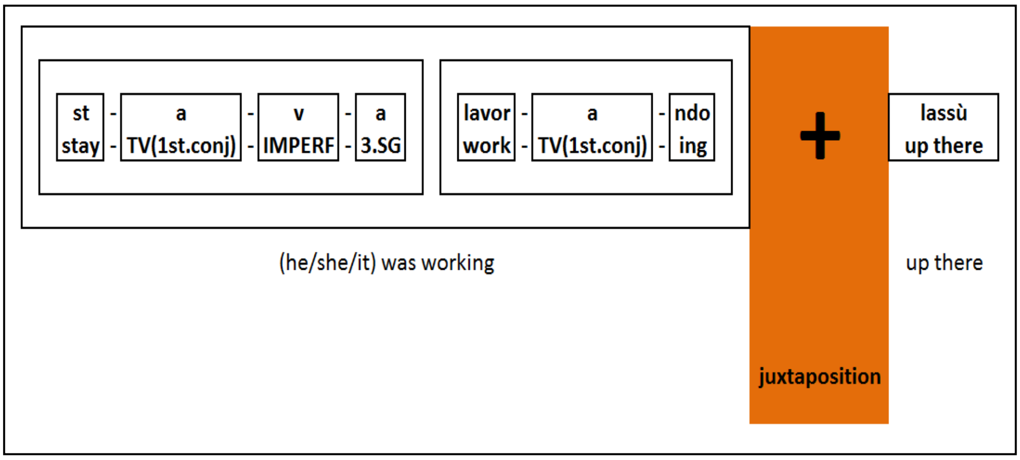


Figure 6.3: the juxtaposition of unit_A and unit_B. The juxtaposition of the two units is coded with a plus (+) sign and is highlighted in red.

The juxtaposition of unit_A and unit_B is allowed only when:

a) Both the sequence AB and the sequence BA are allowed grammatical sequences.

b) The element juxtaposed is neither a compulsory argument of the verb nor a DO.

c) There is no agreement involved between the two elements that are juxtaposed[40].

[40] Requirement (c) was relaxed in the case of seven sentences presenting a dislocated object-NP co-indexed with a resumptive element. Such sentences are discussed in section 8.4.4.1.

Superimposition is the process whereby a fully or partially concrete component unit is used to fill (or elaborate) the slot of a schematic unit. Target sentence SA087 (30) is derived by superimposing unit_B over the slot of unit_A (see fig. 6.4).

30. **CHI: ho preso questi*
ho *preso* *quest-i*
have(PRS.1.SG) taken this-M.PL
"I took these."
(week6.2014.02.18.A: line 756)

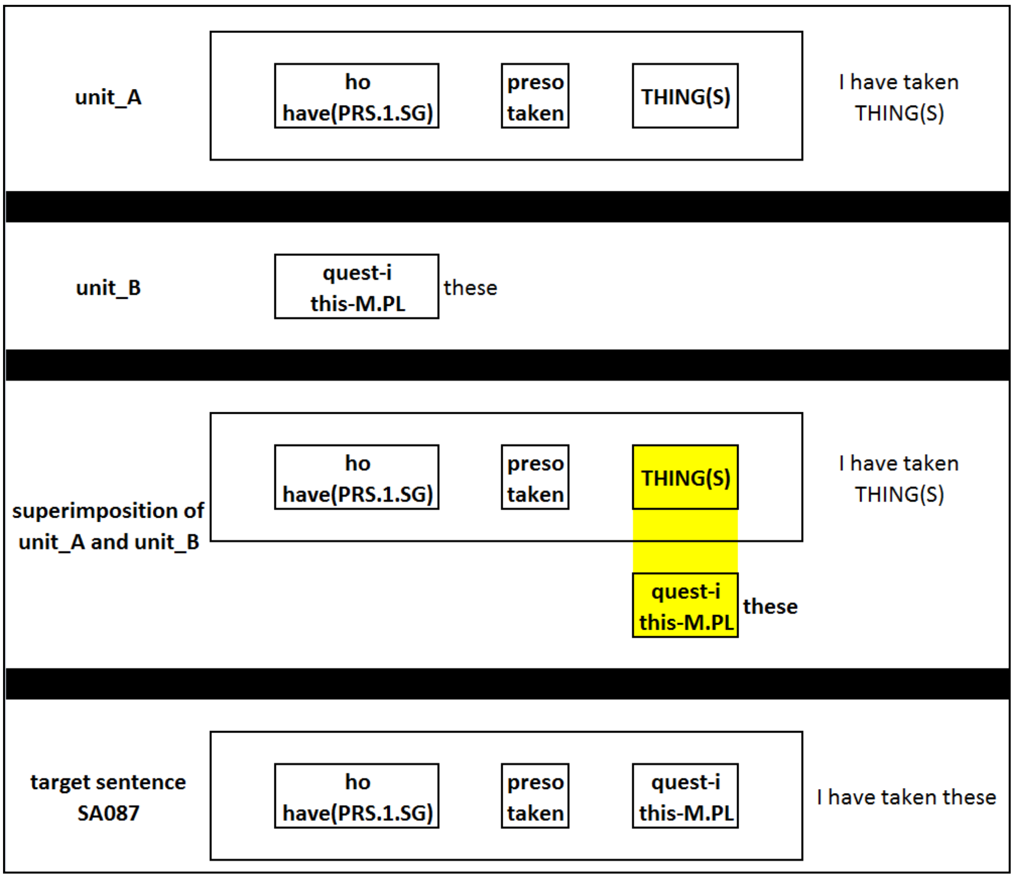


Figure 6.4: illustrating the operation of superimposition. Slot elaboration is highlighted in yellow.

For superimposition to be applied successfully, the filler must correspond to the profile specified by the SLOT. In fig. 6.4, *questi* "these" profiles a *THING* in Langacker's (1987, 2008) terms as does the slot. The slot can be instantiated by any *THING* that can be taken. *Questi* "these" is semantically compatible with a patient role and hence it can elaborate the slot of unit_A.

It is important to point out that superimposition is often a two-way route, particularly when two schemas are superimposed. Target sentence SB108 (31) is derived by superimposing the units in fig. 6.5 and fig. 6.6: each schema elaborates the slot of the other (fig. 6.7). The tense-person-number marker attested in fig. 6.6 elaborates the SLOT of the schema in fig. 6.5. Simultaneously, the root of the verb *portare* “to take” (fig. 6.5) elaborates the *PROCESS* slot of the schema in fig. 6.6, as illustrated in fig. 6.7. Target sentence SB108 (31) is then derived by the **mutual superimposition** (refer back to 2.4.2.2) of two schematic units.

31. TARGET SENTENCE (SB108)
    *CHI: *lo porterà via.*
    l-o *port-er-à* *via*
    clitic.3.ACC-M.SG take-FUT-3.SG away
    “(S/he/it) will take it/him away.”
    (week6.2014.02.18.B: line 897)

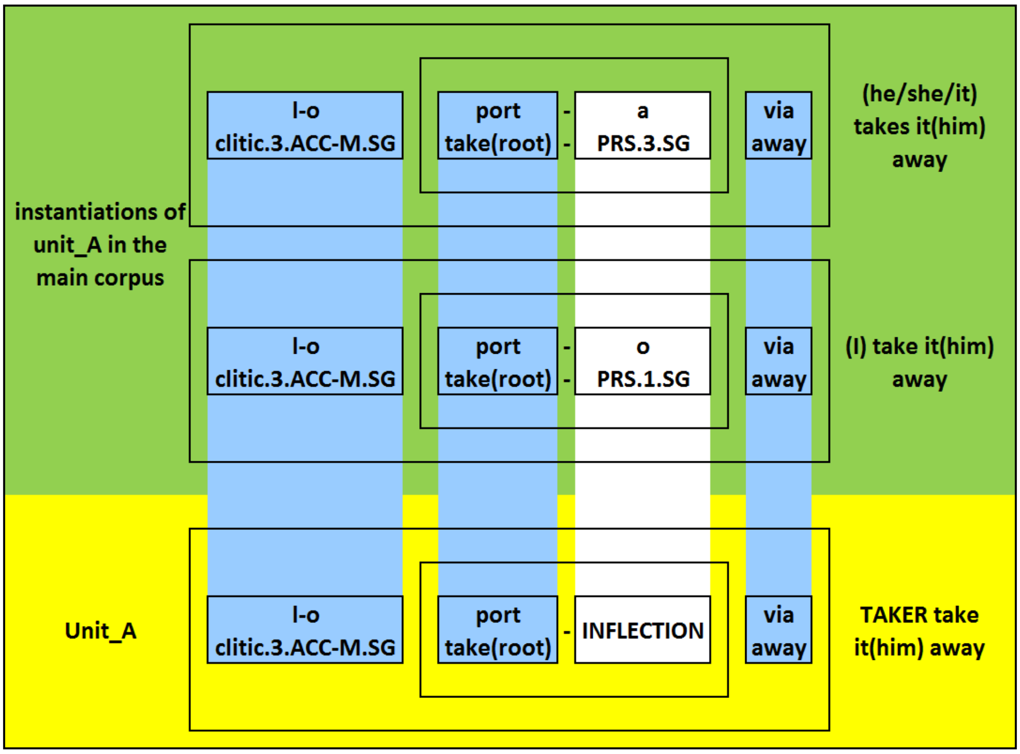


Figure 6.5: Component unit_A of target sentence SB108 (31) (yellow strip) and its instantiations in the main corpus (green strips). Slot formation is highlighted in white. Shared and fixed lexical material is highlighted in blue.

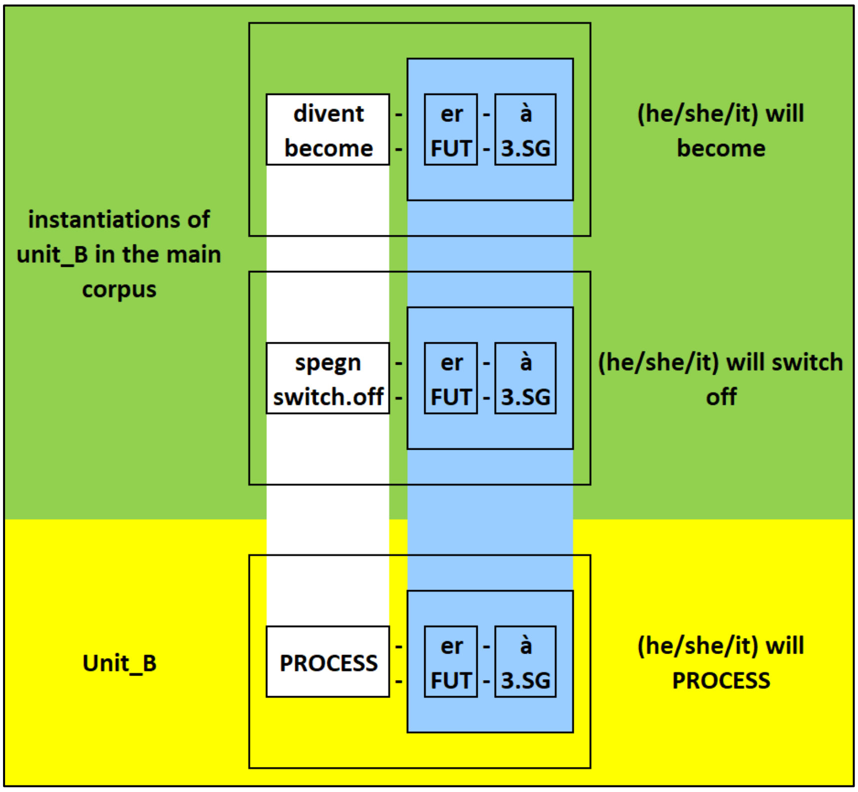


Figure 6.6: Component unit_B of target sentence SB108 (31) (yellow strip) and its instantiations in the main corpus (green strips). See previous figures on how to interpret how colours are used.

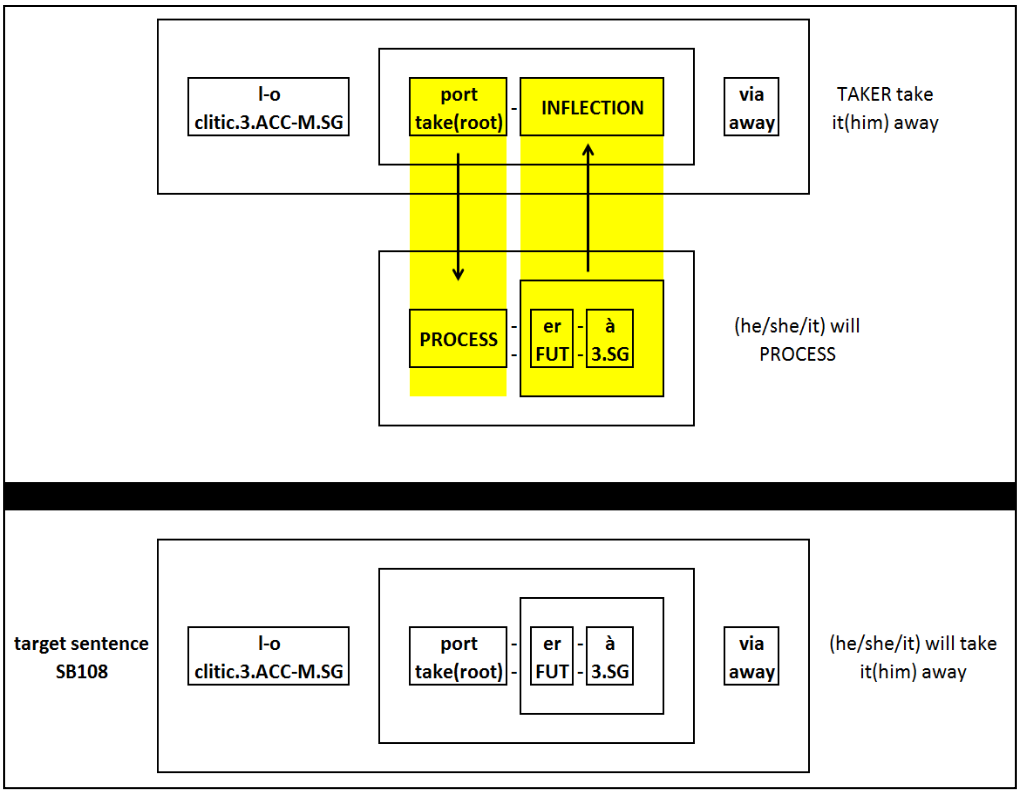


Figure 6.7: deriving target sentence SB108 (31) through mutual superimposition of the schema fig. 6.7 (unit_B) and the schema fig. 6.6 (unit_A). Arrows move from the filler to the elaborated slot.

Mutual superimposition is often found when morphological endings are involved (refer back to 2.2.7 and 2.4.2). Aspect-mood-tense-person-number markers are bound morphemes. It follows that the ability to single out a particular form, such as the ending *–er-à* "-FUT-3.SG", implies the ability to generalise from the various instantiations of the two morphemes (*-er-* and *–à*) when they appear in the ambient language as part of fully formed words (verbs; see fig. 6.6). Aspect-mood-tense-person-number inflections are always the concrete part of word-level schemas (see Booij, 2010) whose slots require to be filled by the root of a particular verb (fig. 6.6). Such slots can be paraphrased as *PROCESS(root)* (refer back to 2.2.7 and 2.4.2).

6.3.6. NUMBER OF OPERATIONS

Each target sentence could be derived by applying zero, one or more operations. A zero-operation derivation is attested when the target sentence is a *Fully Lexically-specific String* which has been uttered twice or more in the main corpus, excluding imitations and self-repeats. Such cases are also called **exact matches** (32-33).

32. TARGET SENTENCE (SA014)
*CHI: ***cosa c' è dentro qua?***
cosa c(i) è dentro qua?
what there.clitic.LOC is inside here
"What's inside here?"
(week6.2014.02.18.A: line 243)

33. COMPONENT UNITS
a) *CHI: ***cosa c' è dentro qua ?***
cosa c(i) è dentro qua?
what there.clitic.LOC is inside here
"What's inside here?"
(week6.2014.02.17: line 161).

b)*CHI: s*ai* ***coa [: cosa] [* p] c' è dentro qua ?***
sa-i ***cosa c(i) è dentro qua?***
know-PRS.2.SG **what there.clit.LOC is inside here**
"Do you know what's inside here?"
(week6.2014.02.17: line 161)

One-operation derivations are those tracebacks that require only one operation (be it superimposition or juxtaposition) to derive a target sentence from its component units (fig. 6.4). When two one-slot schemas are mutually superimposed (fig. 6.7), this also counts as a one-operation derivation, as long as both units are one-slot schemas. The process of derivation is hypothetically without limits. Hence, a given target sentence can be derived by applying any number of operations (either superimposition and/or juxtaposition), as long as the component units meet the frequency threshold. Target sentence SA146 (34) is a two-operation derivation (fig. 6.8).

*34. *CHI: l'ho lasciato a casa della nonna*

l'	*ho*	*lasci-a-t-o*
clitic.3.ACC	have(PRS.1.SG)	leave-TV(conj.I)-PTCP-M.SG

a	*cas-a*	*de=ll-a*	*nonn-a*
to	home-F.SG	of=the-F.SG	grandma-F.SG

"I've left it at grandma's."
(week6.2014.02.14.A: line 1175).

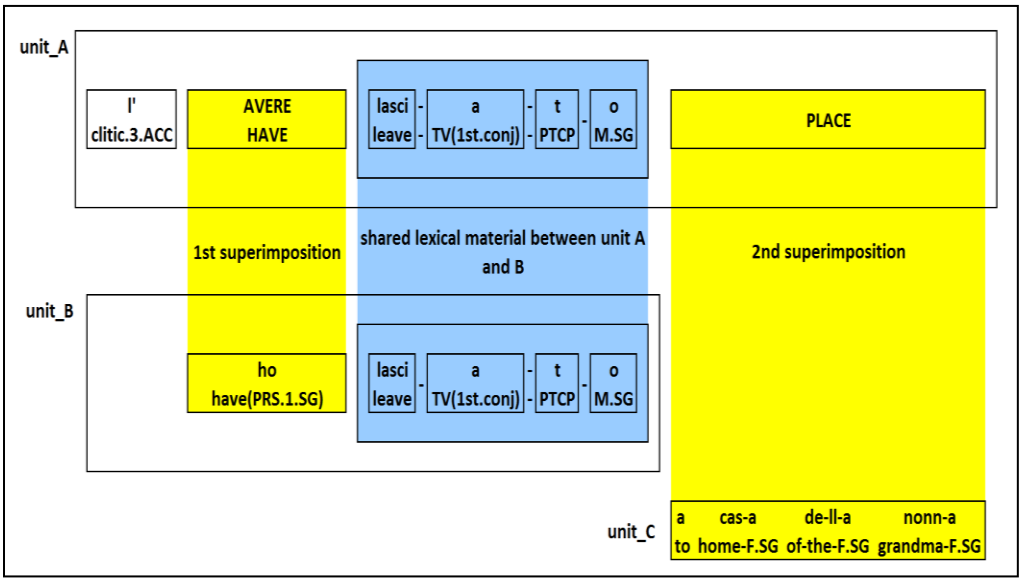


Figure 6.8: deriving target sentence SA146 (34) through two superimpositions (in yellow). The lexically-specific material shared by the two units is highlighted in blue.

6.4. METHODOLOGICAL CHOICES

Now that the method has been fleshed out, the following sub-sections walk the reader through (some of) the methodological choices that yielded the final design.

6.4.1. OTHER LANGUAGES, OTHER SCHEMAS

There are two methodological choices that differentiate this study from both Dąbrowska and Lieven's (2005) and Lieven et al.'s (2009) studies; the way component units are handled and the morpho-syntactic nature of schemas' slots. Such methodological differences are by-products of the need to design a methodology able to realistically account for the morpho-syntactic behaviour of Italian.

6.4.1.1. Word Order criterion

Lieven et al. (2009) considered as schemas those utterances that

> "matched the novel (i.e. target) utterance in the same way, with variation in the **same position** [bold added]".
>
> (Lieven et al, 2009, p. 486)

Similarly, *Fully Lexically-specific Strings* (which they called fixed strings) were strings which presented the exact same words and the same internal order as their target sentences. This is a very sensible choice if English is the language to be investigated as it presents a fairly rigid order of constituents. Italian does not have such rigidity and therefore the internal order of its component units is not rigid either.

Language, as previously pointed out, is an inventory of constructions and speakers draw on their own inventories in order to convey their communicative intentions. Constructions can be thought of as the units speakers use and, if needed, manipulate in order to produce new utterances. Each construction is a form-meaning pairing which maps onto some kind of life event/speech act (i.e. it has a meaning and/or a communicative function).

Thus, an English-speaking child might draw on the fully lexically-specific construction in (35) in order to describe/comment on his/her mother dancing.

35. Mum is dancing.

An important point that needs to be stressed is that such a fixed-phrase is analysed into its component units. Indeed, if one assumes, as UB researchers do, that fully schematic constructions develop from partially schematic ones and these in turn stem from fully lexically-specific constructions, positing that the latter are analysed into their components is essential. It is only by analysing a fully-concrete construction into its components that a child can start developing it into a partially schematic construction (i.e. a lexically-based schema) with one or more slot. More concretely, a child must be able to analyse (35) into at least *mum* and *is dancing*; such an ability is indispensable for the child to start analogising across different occurrences of similar expressions (*daddy is dancing*, *the dog is dancing* and so forth) and developing partially schematic constructions such as *DANCER-is-dancing*.

Even Italian-speaking children supposedly start their journey towards adultlike linguistic competence from fully concrete strings. Those strings, however, just do not happen to have a fixed WO.

Dąbrowska (2014) gives the name **packets** to the units on which speakers of free WO languages draw. Such packets:

a) can be differently schematic: from fully schematic, to partially schematic (Schema-with-Slots) to fully lexically-specific

b) have a certain number of elements (be they words, morphemes or slots), but their order is either not specified or only partially specified.

An Italian-speaking child might rely on the Italian counterpart of (35), in order to describe/comment on her/his mother dancing (36).

36. a) La mamma sta ballando

l-a	*mamm-a*	*st-a*	*ball-a-ndo*
the-F.SG	mum-F.SG	stay-PRS.3.SG	dance-TV(conj.I)-ing

b) *Sta ballando, la mamma*

st-a	*ball-a-ndo*	*l-a*	*mamm-a*
stay-PRS.3.SG	dance-TV(conj.I)-ing	the-F.SG	mum-F.SG

"Mum is dancing."

Thus, if an Italian-speaking child analysed (36) into its components, s/he would be able to infer that *la mamma* "the mum" refers to his/her mum and that *sta ballando* "is dancing" maps onto the action of dancing. As a result, s/he can infer that both (36a) and (36b) are possible.

Both English (35) and Italian (36) are *Fully Lexically-specific Strings* which constitute form-meaning pairings that map onto a meaning, i.e. they are constructions (hence units available for production). Since Italian does not have a rigid WO, the lexically-specific units available to Italian-speaking children can have a flexible WO.

This difference is expressed by the terminology adopted. Studies on English call them fixed strings (Lieven et al. 2009) or fixed phrases (Dąbrowska & Lieven, 2005). Such units are here called *Fully Lexically-specific Strings*, which can have either a fixed (**fixed-strings**) or a flexible (**fully-specific-packets**) internal ordering of constituents.

It follows that this WO criterion be applied to self-repeats and imitations as well. For example, if an Italian child hears (37a) and immediately after utters (37b), it is a case of imitation.

37. a) *MOT: *lo diamo a Roberto*

l-o	*d-iamo*	*a*	*Roberto*
clitic.3.ACC-M.SG	give-PRS.1PL	to	Roberto

"We give it to Roberto."

b) *CHI: *a Roberto lo diamo*

a	*Roberto*	*l-o*	*d-iamo*
to	Roberto	clitic.3.ACC-M.SG	give-PRS.1.PL

"We give it to Roberto."

(fictitious examples)

The child hears a packet of words which forms a coherent unit and describes a specific event; namely, the act of giving X to Roberto. Afterwards, s/he (re)uses it.

As Dąbrowska (2014) points out, the linear order of the constituents of a given packet will depend on various factors such as processing constraints (e.g. memory limitations), information structure, etc. For example, the child might utter (37b) because to him/her the salient element is the receiver. Alternatively, *a Roberto* "to Roberto" is the last element uttered in (37a) and hence the fresher in memory and thus easier to retrieve.

In the same way as some fully-specific units can be thought of as packets of words whose elements are fully phonologically specified, some schemas can be thought of as packets of words which contain one or more schematic elements (slots). Slots are phonologically unspecified units that correspond to specific semantic (*THING*, *PROCESS*, *PLACE* and so forth) and/or morphological entities. They exist in that they are generalisations inferred from their instantiations (also called **fillers**). Fillers are concrete, fully phonologically-specified expressions that can "fill" the slot by virtue of semantic and/or morphological features they share with that particular slot. If a child works out that *daddy, mum* and *the dog* are instantiations of the slot *DANCER*, s/he is also likely to conclude that in Italian the slot can take different positions in the schema. Thus, Italian schemas do not have the syntactic rigidity of English schemas simply because Italian does not have such WO rigidity.

The traceback method designed in this study allows units (be they *Schemas-with-Slots* or *Fully Lexically-specific Strings*) to have a flexible internal WO. Such flexibility is allowed as long as the main corpus provides evidence of it. As a result, units which appear in the main corpus with at least two different linear orders are assumed to be units whose constituents can take different positions (**fully-specific-packets and schematic-packets**). Units whose constituents present only a specific linear order are considered units whose internal order is fixed (**fixed-strings and fixed-schemas)**.

6.4.1.2. A morpho-syntactic approach

Another methodological aspect which differentiates this research from previous studies on English-speaking children's language is that, in this study, schematic constructions can incorporate slots that represent generalisations of verbs' aspect-mood-tense-person-number markers (refer back to fig. 6.5) and nominal affixes (gender and number markers, diminutives and so forth). As a result, bound morphemes can act as both fillers (fig. 6.5) and fixed elements around which schemas are built (fig. 6.6).

Even such a methodological choice, like the previously described WO criterion, stems from the need to have a traceback methodology which can realistically account for how the language under scrutiny is acquired.

Italian speakers' utterances are subjectless 70% of the time (Lorusso et al., 2005) and basically always so when it comes to first and second persons (Benincà with al., 2001). English-speaking children seem to develop schemas with slots in preverbal position (*EATER-eat*); this allows them to be flexible on who performs a particular action. In Italian, such flexibility is (overwhelmingly) obtained by means of verbal inflections[41]. Both subject slots in English and morphological slots in Italian map onto a specific semantic role; namely, agent. Thus, the

[41] Look back at the *INFLECTION* slot of fig. 6.5.

schemas that are to be indentified in this study are morpho-syntactic, rather than syntactic in nature, as Italian itself cannot be accounted for by keeping the two apart.

Additionally, Italian has a fairly complex nominal morphology. Affixes are used to convey information such as gender, number and size (refer back to 1.1.1.2).

For example, Italian has a diminutive affix (-*in*-), which broadly corresponds to the English adjective *little*. Italian can express the size of an entity both analytically and synthetically. English achieves this only analytically.

Hence, in order to describe a small pig, Italian children could use either the affix –*in*- (synthetical means, 38a) or combine the noun with an adjective which agrees in gender and number with it (analytical means, 38b-c).

38. a) *Il maialino*
il *maial-in-o*
the(M.SG) pig-little(DIM)-M.SG

b) *Il piccolo maiale*
il *piccol-o* *maial-e*
the(M.SG) little-M.SG pig(M.)-SG

c) *Il maiale piccolo*
il *maial-e* *piccol-o*
the(M.SG) pig(M.)-SG little-M.SG

"The little pig."

An English-speaking child might rely on (39) in order to speak about a little thing.

39. the little THING.

An Italian-speaking child could rely on the (attested) schema in (40) to express the same concept.

40. Il THING-in-o
il *THING-in-o*
the(M.SG) THING-little(DIM)-M.SG
"The little THING."

In (40) the invariable recurring fixed elements are the article *il* "the(M.SG)", the infix -*in*- and the gender-number marker -*o*. Like the English schema, the Italian schema incorporates fixed information about the size (*little*) of an unspecified entity (*THING*). Additionally, the Italian schema incorporates gender and number information (the article *il* and the final vowel *-o*, both M.SG). The English schema is built around an article and an adjective, the Italian schema around an article (*il*), a gender-number marker (*-o*) and a morphological affix (-*in*-).

The very nature of Italian requires that bound morphemes (whether they belong to verbal or nominal morphology) be indeed considered as both potential fillers that elaborate schemas' slots and potential concrete, recurring lexical material around which schemas are built.

6.4.2. SAMPLING METHOD

6.4.2.1. Sampling Regime

The traceback method requires that each target sentence (e.g. *that's a car*) attested in the **test corpus** be traced back to putative component units attested in the **main corpus**. Such component units are strings (*that's a jeep*, *that's a jar*) that instantiate the same schema (*that's-a-THING*) as the target sentence (*that's a car*). That is, both the main corpus strings and the test corpus target sentence are instantiations of the same construction type. Thus, in order to have a precise insight into the child's language, it is of vital importance to capture on tape as many constructions as possible.

The traditional sampling regime adopted by most longitudinal studies normally involves recording children's spontaneous production for 30 to 60 minutes a week. This captures between 1% to 1.5% of a child's average linguistic experience (Tomasello & Stahl, 2004; Lieven & Behrens, 2012).

Using the Poisson Distribution, Tomasello and Stahl (2004) calculated the different *hit rates* of different sampling regimes, i.e. the probability of a specific sampling regime (e.g. one hour a week) capturing at least one occurrence of different constructions with varying utterance frequencies during one week of recording. For instance, if a given construction is uttered roughly fourteen times a week, a sample as dense as one hour a week has a 20% chance of capturing at least one occurrence of such a construction (hit rate 0.2). It follows that if a construction with similar frequency is found in the **test corpus**, it would be very unlikely to find another instance of the same construction in the **main corpus**.

This might (mis)lead researchers to conclude that the child is not relying on lexically-specific units in order to utter it (and consequently that s/he is somehow relying on more schematic units). However, it might well be that since there is only a 20% chance of capturing that construction on tape, the child has in fact encountered other instances of it.

Previous traceback studies (Lieven et al., 2003, 2009; Dąbrowska & Lieven, 2005) recorded children for five hours a week, which captures 5% to 10% of what children say and hear (Tomasello & Stahl, 2004; Lieven & Behrens, 2012). In the case of a construction that is as frequent as fourteen times a week, such a denser sampling has a 60% chance of capturing on tape at least one occurrence of such a construction (hit rate 0.6). A five-hour-a-week sampling regime therefore captures at least three times as much of children's linguistic experience as traditional samplings (hit rates 0.6 vs. 0.2; Tomasello & Stahl, 2004).

Recording children's production more intensively can therefore help to take a more accurate picture of their language. Consequently, Roberto's spontaneous production was recorded for five hours a week for six weeks in a row.

6.4.2.2. Length of the Study

Item-based constructions make their first appearance when children turn two years old (Tomasello, 2003; 2006a). By the time they are four years old, most English-speaking children have developed a more adultlike linguistic representation of many constructions (e.g. the transitive construction; see Akhtar, 1999).

Ideally then, one should record a child for five hours a week from two to four years of age. Realistically, this is not easily feasible for two main reasons. Firstly, it would need a very deep commitment from the child's family (Lieven & Behrens, 2012), even more difficult to obtain when participants are, as they indeed were in this research, volunteers. Secondly, the interaction between the very nature of this research (a PhD project) and the time needed to transcribe one hour of recording makes it unrealistic. Indeed, transcribing one hour of spontaneous production took up to fifty hours of work, which means that transcribing two years of recordings could have taken up to 26000 hours[42].

A compromise is to record the child so intensely for six weeks in a row and use the last hours (two to four) of recording as a test corpus. The sentences uttered in the **test corpus** represent a picture of the child's language at a given point in development; the given point in development being when recording of the **test corpus** took place (when the child is 2;2.26 in this case). By analysing those sentences, researchers can have a meaningful insight into whether the language spoken at, for example, 2;2.26 can be accounted for in terms of lexically-specific units previously encountered (i.e., attested in the **main corpus**, which represents the child's previous linguistic experience).

6.4.3. FREQUENCY TRESHOLD

The next methodological choice to be discussed is the threshold established in order to determine whether an expression was classified as an available unit that the child could retrieve in order to express his communicative intentions. Hence, whether an expression was part of his inventory of constructions (i.e. his grammar).

In Lieven et al. (2009) an expression was given status of available unit if the child uttered it at least once in the main corpus, excluding imitations and self-repeats. Although it might be realistic to assume that an expression that has potentially been uttered up to ten times[43] is a unit available to the child, there are two aspects of such a methodological choice that are problematic.

Firstly, strings other speakers utter cannot be classified as available units. An important assumption of UB researchers is that language is learnt from the input. Consequently, it would be sensible to assume that, if an expression is heard

[42] 52 weeks per year means 104 (52 x 2) weeks of transcriptions. Each week consists of 5 hours of recordings, meaning a total of 520 (104 x 5) hours. Each hour could have taken up to 50 hours of transcribing work, meaning that transcriptions could have taken up to 26000 (520 x 50) hours. If one worked up to 50 hours a week, transcribing could take up to 520 (26000 : 50) weeks; that is, about 10 years (520 : 52 = 10).

[43] If the sampling regime adopted (5-6 hours per week) captures 5% to10% of a child's average linguistic experience, an expression which is uttered once could have in fact been uttered up to 10 times (10:100 = 1:X).

several times, it must be entrenched enough to be used in production. Ideally then, input should be considered as a source of component units.

Secondly, if anything that the child has said once were classified as an available unit, it would be impossible to distinguish between fully-specific and schematic units. If the target sentence were (41) and its only precedent were (42), it would be assumed that the child drew on the schema in (43).

41. I eat an apple.

42. I eat pasta.

43. I-eat-THING_EATEN.

The problematic part of such a methodological choice is that, since (42) is the only precedent of (41), it is not possible to establish whether the utterance in the main corpus is indeed an instance of the schema in (43), or a fixed-string. In the former case the derivation is possible, in the latter it is not.

As for this study, considering strings that appeared in the main corpus only once as available units would have a further consequence; it would not be possible to disambiguate between units whose linear order is fixed and packets.

Dąbrowska and Lieven (2005) established that an expression was an available unit if it had been uttered by any speaker at least twice in the main corpus, excluding imitations and self-repeats. If a child utters an expression twice (excluding imitations and self-repeats), this is indeed evidence that such an expression is part of his/her inventory of constructions, as it could have been uttered up to twenty times (10:100 = 2: X). Similarly, if an expression is potentially heard up to twenty times, it should present enough type and token frequency to be considered as an available unit.

Raising the threshold to two occurrences, irrespective of who utters the particular sentence at stake, seems to be more sensible with respect to two important issues. Firstly, it acknowledges the importance of the input and that language is learnt from it. Before being able to use a particular construction in production, the child must have heard it. Secondly, it makes it possible (at least in principle) to disambiguate between fully-specific and schematic units on the one hand, and between packets and units with a fixed WO on the other.

Thus, the frequency threshold for strings to be considered as instantiations of available units is two occurrences (excluding imitations and self-repeat), irrespective of who uttered them.

6.4.4. ACCOUNTING FOR AGREEMENT

A method that aims to enquire into the acquisition of a morphologically rich language such as Italian must be able to account for the various types of agreements that are often found in Italian sentences. The following sub-sections illustrate and discuss two main issues that are related to agreement and how these were handled.

6.4.4.1. THING, THINGS and THING(S) slots

A *THING* slot was attested any time that a schema contained generalisations over different kinds of NPs, which overwhelmingly had either object or subject function in the schema. Fig. 6.9 shows a schema which presents two *THING* slots (object and subject).

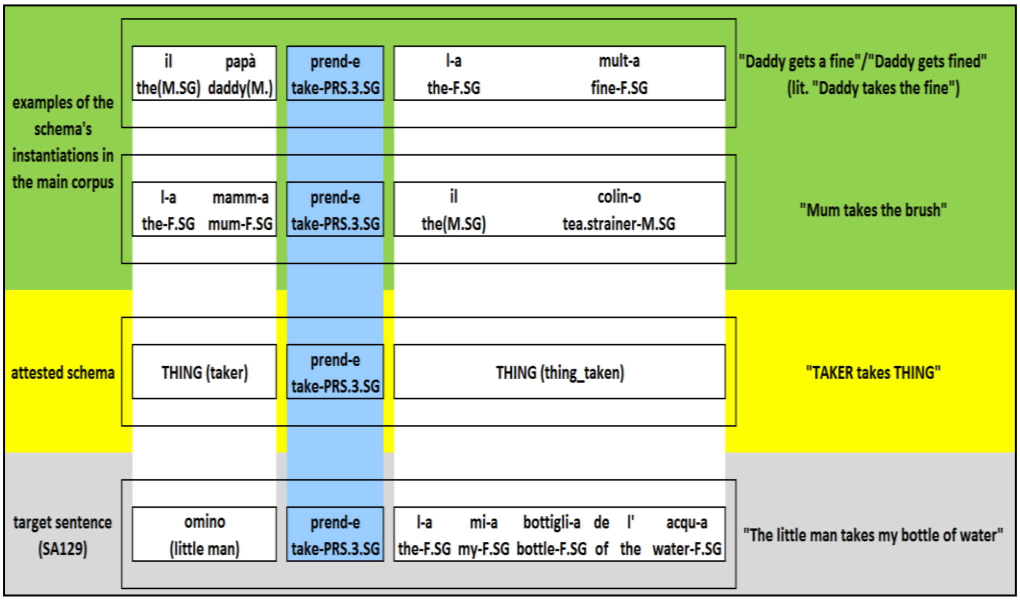


Figure 6.9: a schema (yellow) with two *THING* slots: *THING prende THING_TAKEN* "THING takes THING_TAKEN". Slot formation is in white. Recurring lexical material is in blue

Such a very schematic *THING* slot raises the question as to the basis on which the child does not fill the slot with a plural NP. Such an issue is not problematic in object position, as object NPs do not agree with the verb. The post-verbal slot could be instantiated by both singular and plural NPs and the sentence would still be grammatical (44).

44. *Mamma prend-e i pomodor-i*

mamm-a	*prend-e*	*i*	*pomodor-i*
mum-F.SG	take-PRS.3.SG	the(M.PL)	tomato-M.PL

"Mum takes the tomatoes."

However, using a plural NP in subject position would yield an ungrammatical sentence, as the verb would no longer agree in number with its subject (45).

45. **Loro prende questo*

loro	*prend-e*	*quest-o*
they(NOM.3.PL)	take-PRS.3.SG	this-M.SG

"*They takes this."

This issue was solved by positing three types of phonologically unspecified *THING* slots: THING, THINGS and THING(S). The latter is a slot that can be instantiated by both plural and singular NPs and is the result of generalisations across different specific patients (*THING(S)_TAKEN*, *THING(S)_WANTED*, and so forth). *THING* indicates a singular entity and can be instantiated only by singular NPs; *THINGS* indicates a plural entity and can be instantiated only by plural NPs. This was allowed on the basis that the child, throughout the study, showed understanding of the distinction between singular and plural entities. For example, he was capable of limited counting, he was able to request two or three chocolate sweets when only one was offered to him and he was able to request one last sweet after he had already eaten some. Hence, there was evidence that the cognitive bases to conceptualise plural and singular entities were developed enough to allow the child to draw generalisations that distinguished between plural (*THINGS*) and singular (*THING*).

6.4.4.2. Agreement and overlapping shared lexical material.

Another agreement issue is gender-number agreement between subject and past participle and/or subject and adjective. In the overwhelming majority of cases this was accounted for by the fact that the *THING* slot was both morphologically-specified and also bound to an article. In fig. 6.10, target sentence SA144 (grey strip) is traced back to a schema (yellow strip) in which the article (*il* "the(M.SG)") and the gender-number marker *–o* "M.SG" of the partially specified THING slot account for agreement (in red): the noun is M.SG and hence agrees in gender and number with the past participle (fig. 6.10).

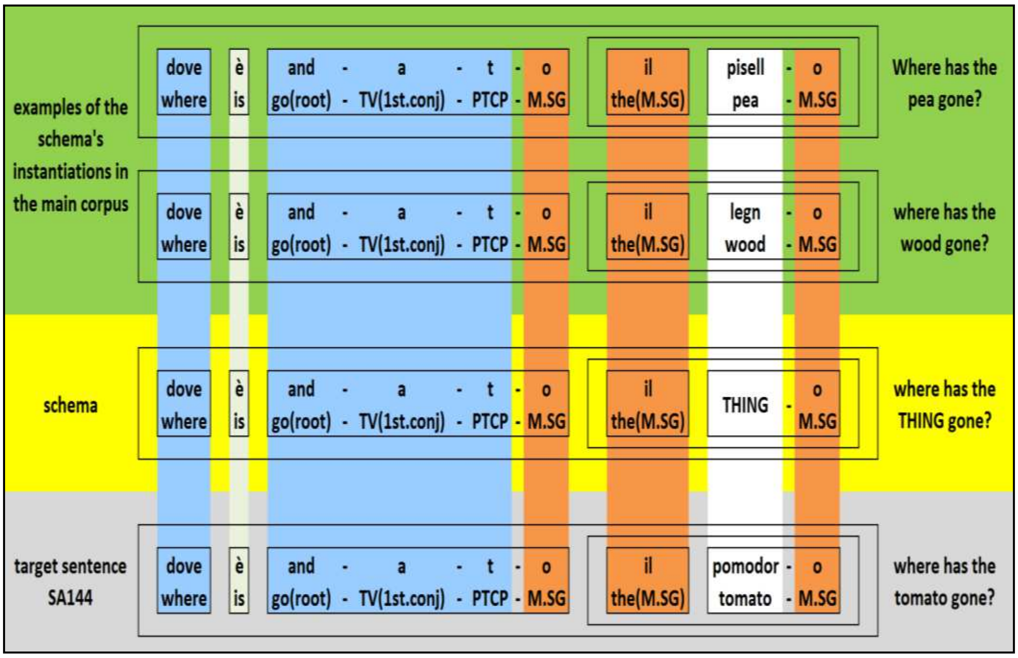


Figure 6.10: Explaining gender and number agreement: the parts highlighted in red show the morphologically-specified, recurring elements in schema (yellow strip), instantiations (green strip) and target sentences (dark grey strip) that account for both gender and number agreement. The light grey highlighting indicates fixed, recurring elements that account for number agreement (namely the PRS.3.SG of the verb *to be*). White parts indicate the slot and its instantiations. Blue parts indicate other shared lexical material, not relevant for agreement.

In some other cases, agreement was accounted for by the article to which the slot *THING* was bound. In the schema in fig. 6.11 (yellow strip) to which target sentence SA003 (fig. 6.11, grey strip) is traced back, agreement between subject and adjective is explained by the article (*l-a*; *the-F.SG*): adjective and article agree in gender and number (in red).

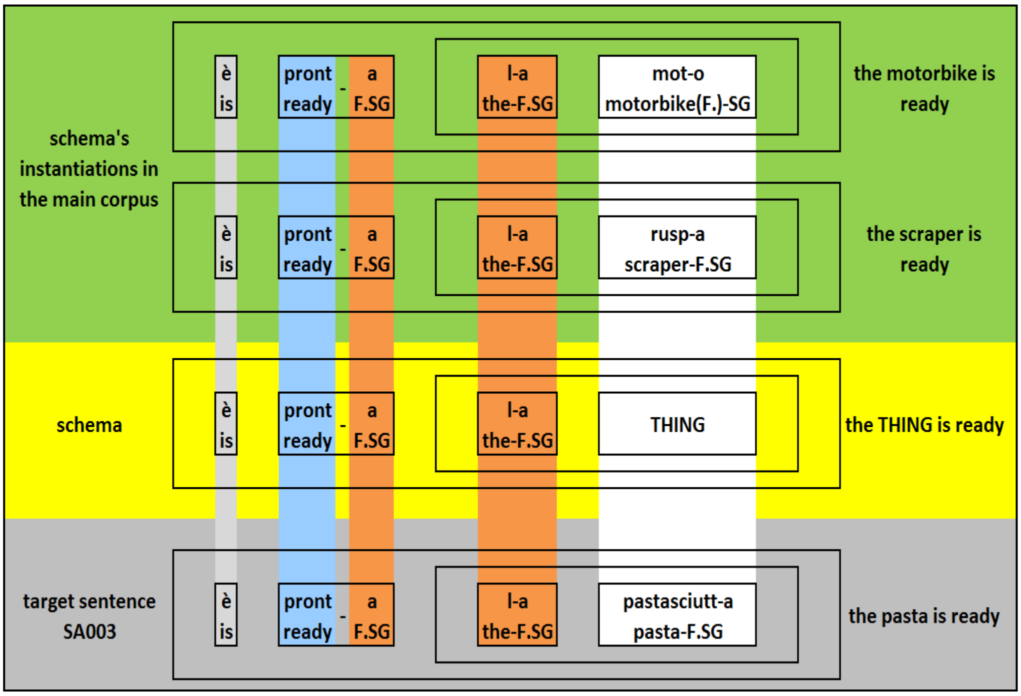


Figure 6.11: see fig. 6.10 on how to read this figure.

The schema in fig. 6.11 (yellow strip) incorporates the feminine definite article (*la*) and hence it (partially) accounts for gender-number agreement between NP and adjective. The slot's profile requires that it be filled with any single entity (*THING* not *THINGS*) that can be ready. However, a unit as large as a single word could be used to fill such a slot. For instance, *pizza* could be ready. However, even *risott-**o*** "rice-**M.SG**" could be ready. Hence, it is not clear what prevents the child from filling the slot of the schema with an ungrammatical masculine NP, yielding a sentence like (46), in which there is no gender agreement between *pront-a* "ready-F.SG" and *l-a* "the-F.SG" on the one side and *risotto-o* "rice-M.SG" on the other.

46. * È *pronta la risotto*
è *pront-**a*** *l-**a*** *risott-**o***
is ready-**F.SG** the-**F.SG** rice-**M.-SG**
"The rice is ready."

The issue of agreement in cases like fig. 6.11 was solved by preferring the largest and most specific units over smaller and more abstract ones. Dąbrowska and Lieven (2005, pp. 460-461) note that the hypothesis that speakers might prefer to activate larger and more specific units might help to account for why children do not normally make many agreement mistakes (refer back to 2.4.2.2, fig.2.12). In particular, shared lexical material between the units superimposed is likely to play a crucial role in pre-empting ungrammatical sentences and is likely to represent a foothold towards the development of correct grammatical behaviour (particularly agreement) (see also Dąbrowska, 2004).

Target sentence SA003 (fig. 6.11, grey strip) is built around the schema *è pronta la THING* "the THING is ready" and has three possible component units (fig. 6.12) and two possible derivations (fig. 6.13). In fig. 6.13a the two units (A and C) are superimposed by virtue of the fact that the filler matches the (semantic) requirements of the slot. In fig. 6.13b, the two units (A and B) are superimposed by virtue of the matching characteristics between slot and filler **and** their shared lexical material (*la*, highlighted in blue).

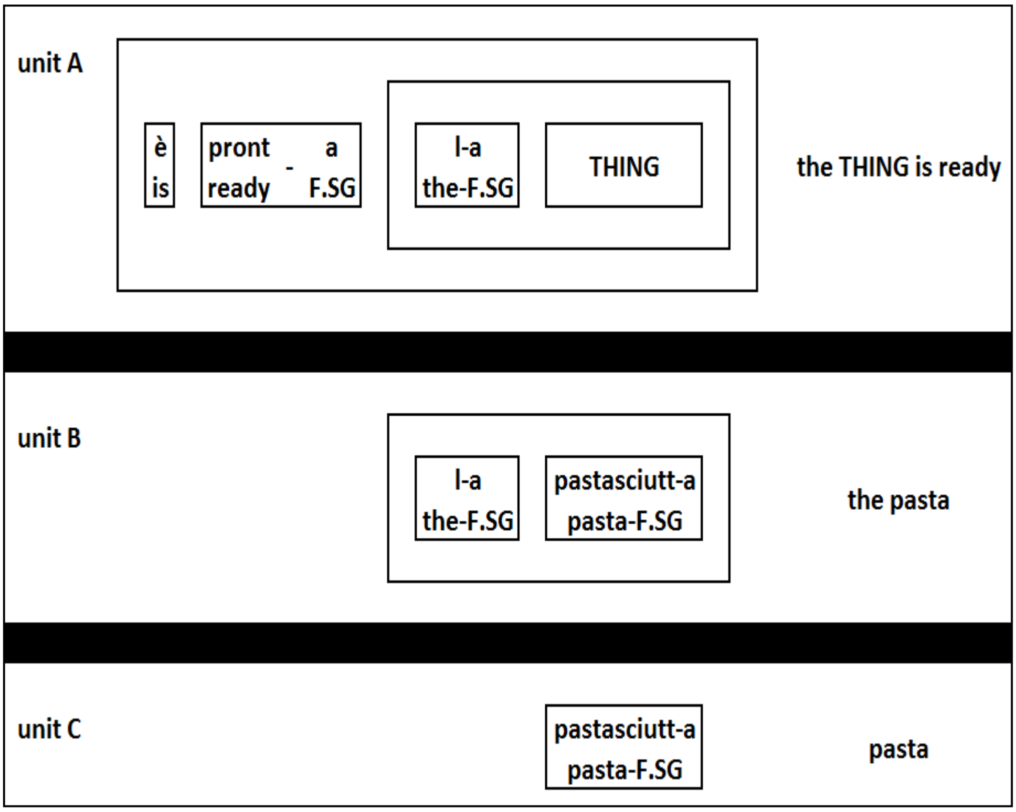


Figure 6.12: Three possible component units of target sentence SA003.

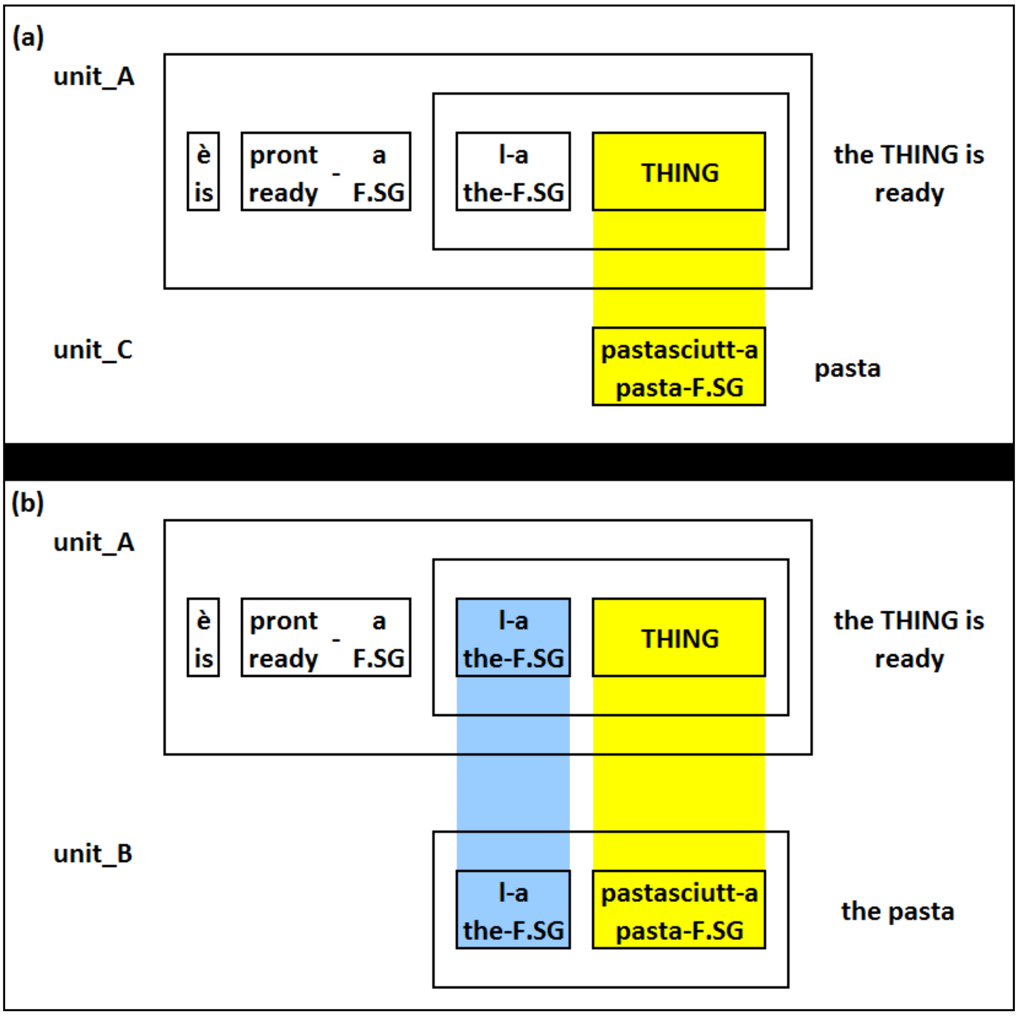


Figure 6.13: Two ways of deriving target sentence SA003 in fig. 6.12. Slot elaboration is highlighted in yellow. Shared lexical material is highlighted in blue.

By assuming that the largest units are chosen over smaller units, it is possible to create a method that accounts for why the child fills the slot of the schema with an appropriate F.SG word and not with an ungrammatical M.SG noun (as in 46). Indeed, in the case of (46), if a unit larger than *risotto* were to be activated, this would likely store the appropriate masculine article that the word requires; either *il* "the(M.SG)" or *un* "a(M.SG)". In the latter case, there would be no correspondence between the article in the schema and the article in the unit that contains the possible filler (fig. 6.14, in green). This lack of correspondence might act as deterrent to the building of ungrammatical sentences. Conversely, shared lexically-specific material between the units activated might help the child to combine the schema with a grammatically appropriate filler (fig. 6.13b).

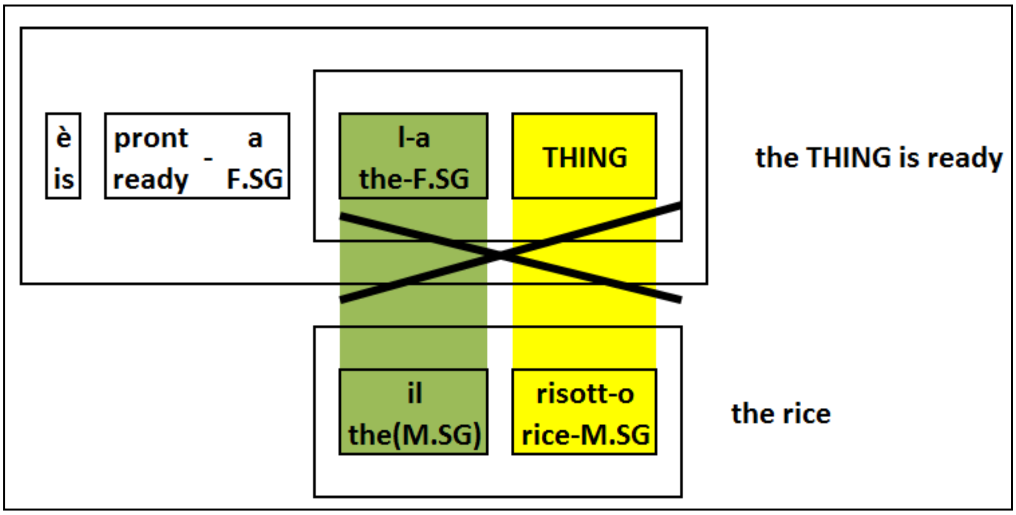


Figure 6.14: activating larger units prevents ungrammatical sentences.

Such a methodological choice (preferring larger and more specific units) is not meant as a claim that correct agreement is due to children activating larger units and that agreement errors are brought about by the activation of smaller units. Nevertheless, it is a design choice that helps to account for gender-number agreement between various elements in Roberto's sentences and it therefore creates a method able to account for his linguistic production.

Hence, in cases like SA003 (fig. 6.11, grey strip), only derivations like 6.13b were considered successful. That is, in order for a derivation to be considered successful, all agreements in the target sentence had to be accounted for. This could be obtained in one of two following ways:

a) the target sentence was traced back to a schema that fully specified all the morphological inflections needed to deliver correct agreement (fig. 6.10)

b) the target sentence was traced back to two units which shared lexically-specific concrete material on the basis of which it was possible to account for the specific (morphological) form the child used (6.13b)[44].

[44] An exception was made for seven sentences involving agreement between an accusative resumptive clitic and a co-indexed object. Such sentences, and the reasons why an exception was made, are discussed in 8.4.4.1.

7.

RESULTS

This chapter reports and discusses the results of the traceback analysis. Firstly (7.1), a purely quantitative overview of the results is given. Section 7.2 provides a more qualitative insight into the results and the kinds of units yielded by the analysis. Finally, 7.3 compares the results with previous findings regarding English-speaking children.

7.1. QUANTITATIVE RESULTS

7.1.1. GENERAL OVERVIEW: COMPARING DIFFERENT METHODS

The method identified 768 sentences to be analysed out of a total of 993 intelligible multi-word sentences (refer back to table 6.3). Tab. 7.1 and figure 7.1 give an overview of the results when different methods are adopted. Method_A is the method adopted for the main analysis and its details have been fleshed out throughout chapter 6. For reasons of space, a detailed description of the other methods is reported in Appendix_II (chapter 20.2).

Table 7.1: Results: using different traceback methods to analyse Roberto's target sentences. Method_A is the method adopted for the main analysis (refer back to chapter 6). For each method, the frequency threshold a precedent had to meet in order to be considered as an available component unit and who could have uttered it are reported in the "method's description". Hence, in Method_C, target sentences were traced back to strings that the child (and the child only) uttered in the main corpus at least once (excluding imitations and self-repeats).

METHOD	method's description	number of operations needed for successful derivations					fails	total no. of utterances
		exact match (0 operation)	1	2	3	4 or more		
Method_A	2 any speaker	113	219	142	90	69	135	768
Method_B	3 any speaker	94	205	135	101	71	162	768
Method_C	1 child only	109	216	131	91	55	166	768
Method_D	1 any speaker	144	252	140	99	59	74	768
METHOD	method's description	type of operation needed						
		sup	juxt	tot				
Method_A	2 any speaker	1062	25	1087				
Method_B	3 any speaker	1091	23	1114				
Method_C	1 child only	984	18	1002				
Method_D	1 any speaker	1069	19	1088				

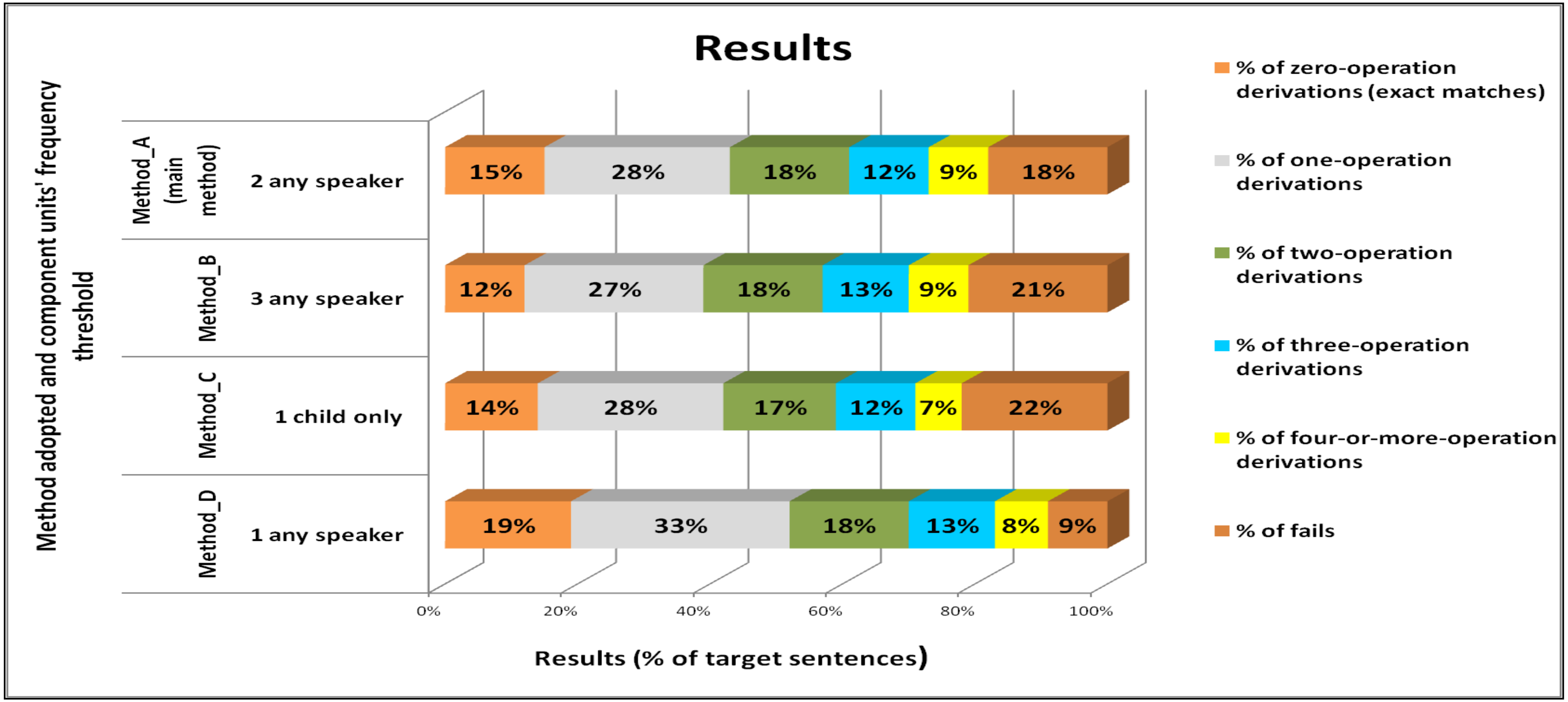


Figure 7.1: Comparing the results of different traceback methods. Method_A is the method adopted for the main analysis (refer back to chapter 6). For each method, the frequency threshold a precedent had to meet in order to be considered as an available component unit and who could have uttered it are reported on the left hand-side of each bar. Hence, in Method_C, target sentences were traced back to strings that the child (and the child only) uttered in the main corpus at least once (excluding imitations and self-repeats).

According to the method adopted by this study (Method_A), 82% of Roberto's target sentences can be accounted for in terms of lexically-specific units extrapolated (learnt) from the input (fig. 7.1, upmost bar). Furthermore, results pertaining to Method_D (fig. 7.1, bottom line), show that about 19% (144) of target sentences are exact matches of strings attested at least once in the main corpus: i.e., they are sentences Roberto has already encountered (sometimes with the same WO, sometimes with a different WO). Method_D also shows that 9% of sentences (74) cannot be accounted at all by what the child has previously said or heard (fig. 7.1, Method_D, bottom bar).

However, it is worth restating that these figures pertain only to the test corpus sentences that are considered novel. When all intelligible multi-word sentences are considered (993 sentences), the proportion of successful derivations under Method_A rises to 86% (table 7.2). If one considers that zero-operation derivations are repetitions of already encountered strings, the proportion of non-novel sentences makes up about 34% of Roberto's entire intelligible production (zero-operation derivations + Imitation and self-repeats (partial and full) + songs, poems and nursey rhymes + instantiations of an already attested target sentence; fig. 7.2).

Table 7.2: results as a proportion of all (993) Roberto's intelligible multi-word sentences using Method_A.

TYPE OF SENTENCE	no.	%	is the sentence novel/creative?
two-and-three-operation derivations	232	23%	yes
one-operation derivations	219	22%	yes
fails	135	14%	yes
zero-operation derivations (exact matches)	113	11%	no
instantiations of an already attested target sentence type	78	8%	no
imitations and self-repeats (full and partial)	77	8%	no
songs, poems and nursey rhymes	70	7%	no
four-or-more operation derivations	69	7%	yes
total	**993**	**100%**	

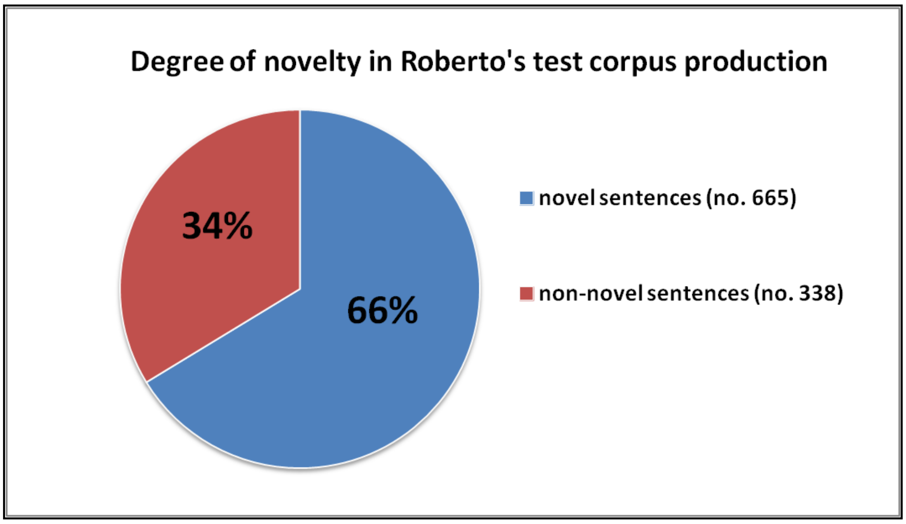


Figure 7.2: degree of novelty in Roberto's test corpus intelligible multi-word sentences.

7.1.2. NUMBER AND TYPES OF OPERATIONS

Under all methods, superimposition is the most used operation and juxtaposition accounts for only about 2% of operations (table 7.1). Hence, Roberto seems overwhelmingly to rely on partially schematic constructions, whose slots are filled by appropriate lexically-specific material.

Looking back at Method_A, when only novel utterances are considered (fig. 7.1, topmost bar), 61% (474) of sentences require zero to two operations to be traced back. 28% of all tracebacks are one-operation derivations and a further 15% are exact matches. Less than 10% require four operations or more.

Target sentence SB180 (47) is the sentence that needed the largest number of operations and is the only ten-operation derivation. Its precedents and its derivation can be found in Appendix_II (ch. 20.2.2).

47. *CHI: <sai> [/] sai che io (.) ho fatto una corsa (..) grande + e sono arrivato (.) dal Luca +. +. e ho fatto pf@o.*
sai che io ho fatto una corsa grande e sono arrivato dal Luca e ho fatto puf.
sa-i *che*
know-PRS.2.SG that
io *ho* *fatto* *un-a* *cors-a* *grand-e*
I have(PRS.1.SG) done a-F.SG run-F.SG big-SG
e *sono* *arriv-a-t-o* *da=l*
and be(PRS.1.SG) arrive-TV(conj.I)-PTCP-M.SG at=the(M.SG)
Luca *e* *ho* *fatto* *puff.*
Luca and have(PRS.1.SG) done puff.
("You know that I have done a big run and I have arrived at the Luca and I've done ʽpuff'?").
"Did you know that I ran a long way and got to Luca and went ʽpuff'?"
(week5.2014.02.18.B: lines 1423 and 1425)

Overall, four-operation derivations represent 65% of all derivations requiring four operations or more and about 6% of the whole dataset (table 7.3).

Table 7.3: Target Sentences that required four or more operations to be derived from their component units.

sentences that required four or more operations to be derived	no.	% of 69 sentences that required four or more operations	% of 768 total target sentences
four-operation derivations	45	65%	6%
five-operation derivations	17	25%	2%
six-operation derivations	4	6%	0.5%
seven-operation derivations	1		
eight-operation derivations	1	4%	0.5%
ten-operation derivations	1		
total	**69**	**100%**	**9%**

7.1.3. PROBLEMATIC UTTERANCES: FAILS

The analysis undertaken following Method_A cannot derive 135 target sentences (18% of the entire corpus). A sentence that cannot be derived is classified as a **fail**. Fails can be of three types: Lexical (38 sentences), Constructional (90 sentences) or both Lexical and Constructional (7 sentences). Table 7.4 reports how fails are distributed across these three types.

Table 7.4: Distribution of fail types.

Method_A: fail type	no. of target sentences that could not be sucessfully derived	% of 768 target sentences	% of 135 fails
constructional	90	12%	67%
lexical	38	5%	28%
both constructional and lexical	7	1%	5%
total	**135**	**18%**	**100%**

7.1.3.1. Lexical Fails

When a target sentence contains a word whose root either does not appear at all in the main corpus or is attested only once, this constitutes a **lexical fail**. Lexical fails represent 28% of sentences which could not be traced back successfully (38 out of 135). For instance, in target sentence SA025 (48) the child uses the word *cer-a "wax-F.SG"*, whose root (*cer-*) is not attested in the main corpus. Neither the plural (*cere*; "waxes") nor the singular (*cera*; "wax") form is found. Hence, even if the putative schema out of which the sentence was supposedly built (49) is attested, the sentence cannot be fully derived.

48. *CHI: *è (..) una cera.*
    è un-a cer-a
    is a-F.SG wax-FG
    "(It) is a wax."
    (week.6.2014.02.18.A: line 304)

49. *È una THING-a*
    è un-a THING-a
    is a-FSG THING-F.SG
    "(It) is a THING."

Overall there are 45 lexical fails; 38 plus 7 that co-occur with constructional fails. Of these 45 sentences, 24 (53%) can be accounted for by Method_D, that is, 53% of lexical fails are sentences containing a word whose root is attested in the main corpus only once. The remaining sentences that are classified as lexical fails (47%) contain words whose roots are not attested at all in the main corpus.

7.1.3.2. Constructional fails

Constructional fails are those sentences that cannot be derived by applying juxtaposition and/or superimposition to precedents that are classified as available units. Such fails represent about 13% of the whole corpus and about

72% of all fails (refer back to table 7.4). Of these, 7 sentences (about 5% of fails and 1% of the whole corpus) contain both a constructional and a lexical fail.

Constructional fails can be grouped in two main categories: constructional fails that can be accounted for by using other methods, such as Method_C and Method_D, and those that cannot be accounted for under any of the methods adopted. I shall refer to the former as **Soft_Constructional_Fails** and to the latter as **Hard_Constructional_Fails**.

Tables 7.5 and fig. 7.3 show that about 52% of constructional fails (50 sentences, about 7% of the whole corpus) are **Hard_Constructional_Fails**, whereas the remaining 48% are **Soft_Constructional_Fails** (47 sentences, about 6% of the whole corpus).

Table 7.5: Constructional fails of Method_A analysed with other methods.

Constructional fails of Method_A analysed with other traceback methods					
Method	method's description	Successful	**lexical	*constructional	total (no. of sentences)
Method_A	2 any speaker	0	0	97	97
Method_C	1 child only	27	7	63	97
Method_D	1 any speaker	44	3	50	97

*sentences that contain a constructional fail or contain a constructional and a lexical fail

**sentences that do not classify as constructional fails but that contain a lexical fail

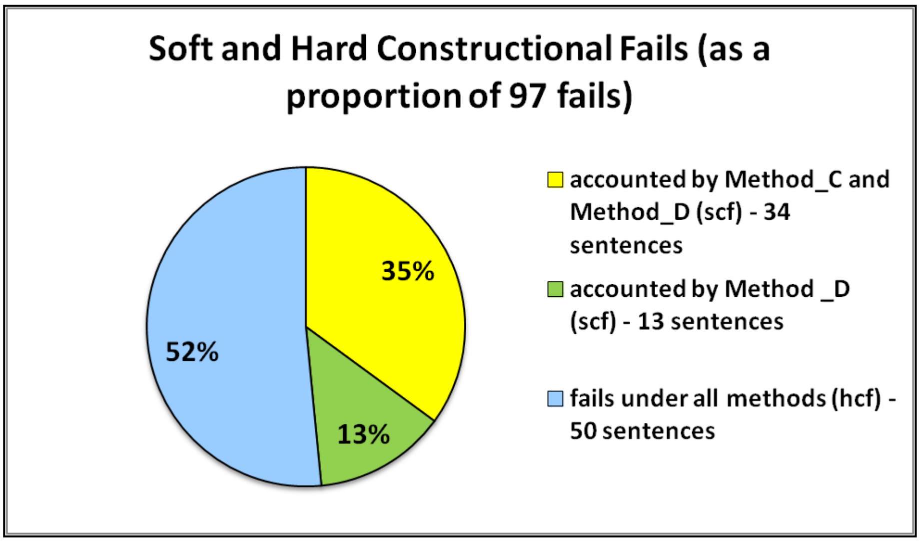


Figure 7.3: Soft (scf) and Hard (hcf) Constructional Fails.

Hard_Constructional_Fails are discussed in detail in section 8.4. As for Soft_Constructional_Fails, they can be grouped into three main sub-categories:

a) Target sentences for which one of the putative component units is attested only once in the main corpus. For example, for target sentence SB312, the only two-slot schema available is instantiated in the main corpus only once in the child's speech (fig. 7.4).

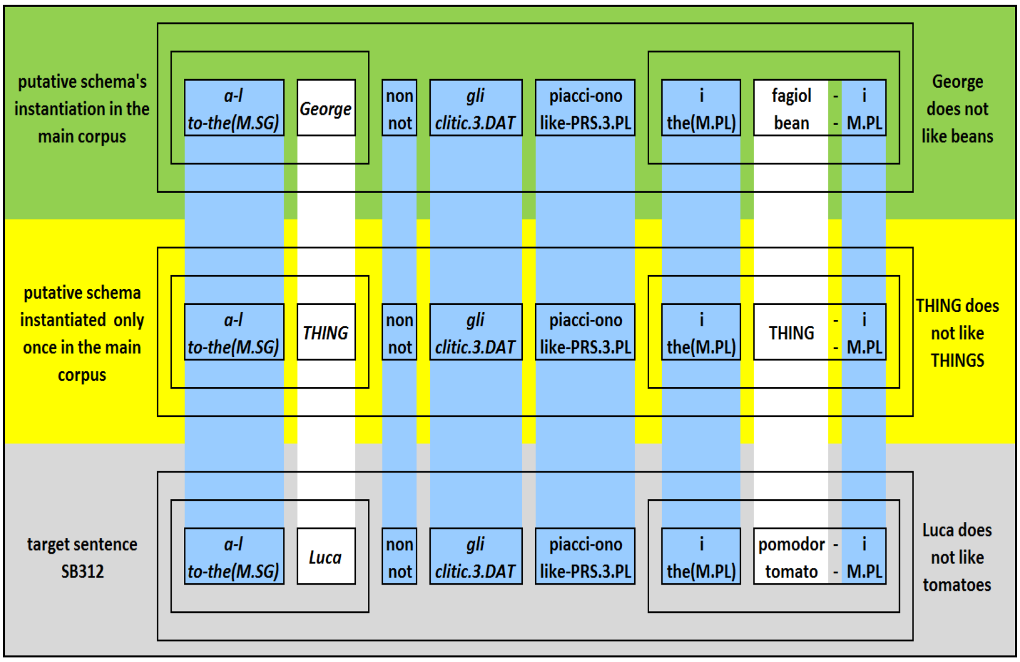


Figure 7.4: Target sentence SB312 (grey strip), its only putative precedent (green strip) and the schema they both instantiate (yellow strip). Slots are in white, whereas shared concrete material is in blue. Elements in *italics* are co-indexed. For the type of construction used in target sentence SB312, refer back to 1.4.3, sentences (42)-(45).

b) Target sentences for which one of the putative (schematic) component units is attested twice or more, but the unit's instantiations in the main corpus present only one (fixed) linear order (e.g. VS), which does not match the ordering used by Roberto in the target sentence (e.g. SV).

c) Target sentences for which one of the putative (schematic) component units is attested twice or more, but its slot is instantiated by only one specific item and hence cannot be classified as a slot.

Type C Soft_Constructional_Fails can be attested at the clause or at the word level. Fig. 7.5 shows a type C Soft_Constructional_Fail at the clause level. The first slot (highlighted in red) of the putative schema is instantiated only by the item *t-i* "clitic.2.SG-DAT", whereas the target sentence contains *m-i* "clitic.1SG-DAT". Since for a slot to be created the same recurring lexical or morphological material must occur at least twice with at least two different fillers, such an element cannot be considered a slot and is treated as a fixed part.

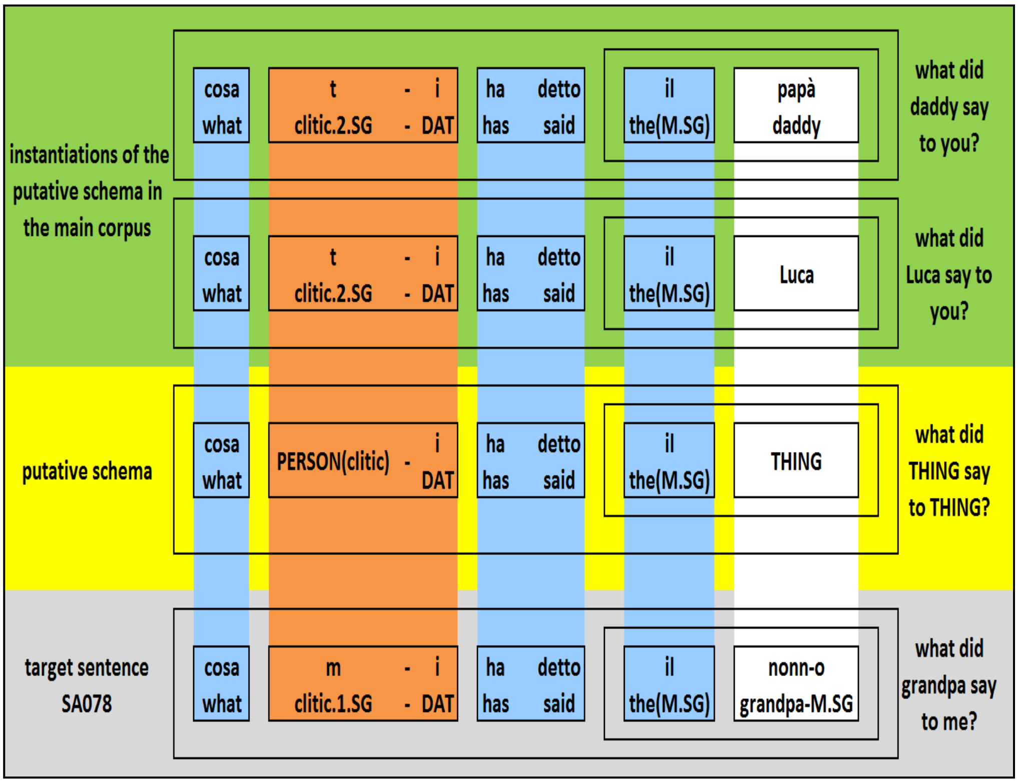


Figure 7.5: Constructional fails at the clause level. Target sentence SA078 (grey strip), its precedents (green strip) and the putative schema they instantiate (yellow strip). Highlighted in red is the putative slot of the schema which does not meet the type variance requirements to be considered as such. Successful slot formation is highlighted in white and recurring lexical material is in blue.

Type C Soft_Constructional_Fails **at the word level** are those sentences containing words that are attested twice or more in the main corpus in one (inflectional) form only and the particular form used by Roberto does not match exactly. For instance, in target sentence SB356 (fig. 7.6, grey strip) the word *pappagall-i* "parrot-**M.PL**" is used. However, only the singular form (*pappagall-o*; "parrot-**M.SG**") is attested. Consequently, the putative schema *pappagall-GENDER.NUMBER* (fig. 7.6, yellow strip) is not classified as an available unit and the sentence cannot be fully traced back. I shall refer to constructional fails at the word level as **morphological fails**, whereas all other Soft_Constructional_Fails discussed in this section and all Hard_Constructional_Fails can be labelled as **syntactic fails**. Table 7.6 shows that morphological fails (15 sentences in total) represent up to 11% of all fails (about 2% of all target sentences).

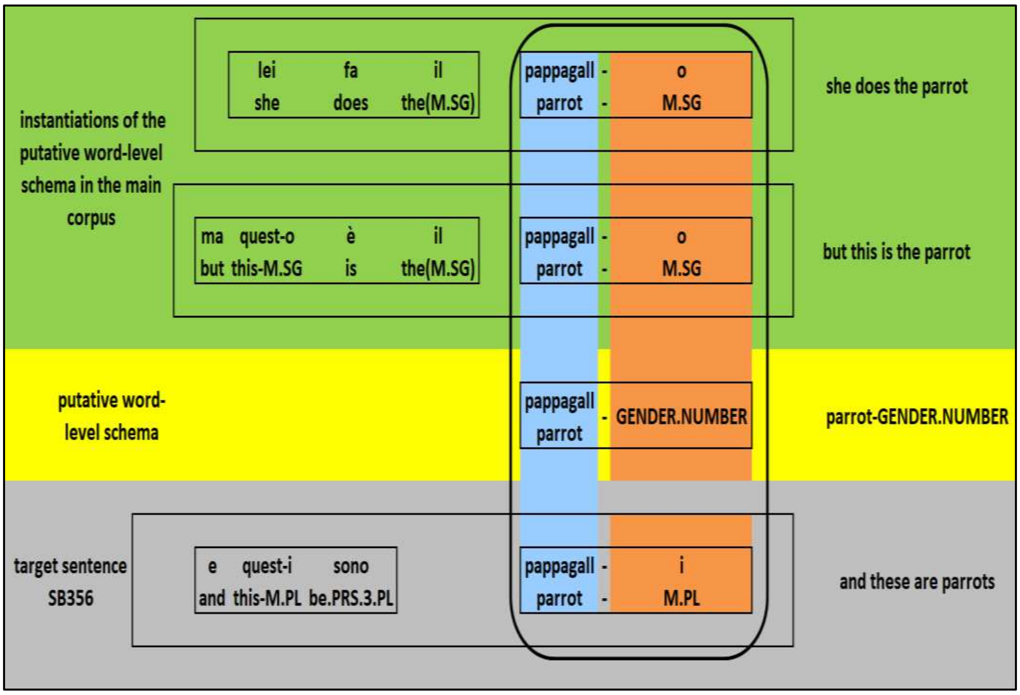


Figure 7.6: Constructional fails at the word level (i.e. MORPHOLOGICAL FAILS). Target sentence SB356 (grey strip). The schema that does not meet the type variance requirement (yellow strip), its instantiations (green strip) and the relevant part of the target sentence are enclosed in the rounded rectangle. Highlighted in red is the putative slot of the schema which does not meet the type variance requirements to be considered as such. The fixed part of the schema is highlighted in blue.

Table 7.6: Syntactic, morphological and lexical fails under Method_A.

fail type	no. of sentences containing at least one fail	% of 135 fails	% of 768 target sentences
syntactic fails	77	57%	10%
lexical fails	38	28%	5%
morphological fails	11	8%	1.5%
both syntactic and lexical fails	5	4%	1%
both syntactic and morphological fails	2	2%	0.1%
both morphological and lexical fails	2	1%	0.1%
total	**135**	**100%**	**17.7%**

7.1.3.3. Fails and the UBA to LA

UBAs posit that children learn their native language by drawing generalisations from the concrete strings they experience. Of the two types of fails identified, **constructional fails** are the most problematic for a UB account of LA, whereas **lexical fails** can be easily accommodated by the theoretical framework adopted. The latter are attested in those sentences in which the child utters a word whose root either does not appear or appears only once in the main corpus. Considering the often arbitrary link between form and meaning of lexical items, it appears safe to claim that if a child utters (and thus knows) a word, it is because s/he has acquired it from language input and not because s/he is relying on some kind of abstract representation, which maps a phonological shape onto its referent. Since the sampling regime adopted by this study captures 5% to 10% of a child's linguistic experience, it is likely the case that the sampling regime simply failed to catch such words on tape (see both Lieven at al. (2009) and Dąbrowska and Lieven (2005) for a similar argumentation).

Conversely, constructional fails (syntactic ones in particular) are more difficult to account for. On the one hand, it is indeed possible that the sampling regime simply failed to record some of the constructions encountered by Roberto. After all, 48% of them (the so called Soft_Constructional_Fails), simply fail to reach the frequency threshold and/or to meet variance requirements. On the other hand, the fact that Roberto uses constructions that cannot be classified as available units could be interpreted as evidence that he is capable of more abstract generalisations and hence that his grammatical competence is not fully lexically-bound.

A more detailed discussion of constructional fails (and how they could be analysed and interpreted) is presented in the next chapter.

7.2. AN INSIGHT INTO ROBERTO'S INVENTORY OF CONSTRUCTIONS

Now that a quantitative overview has been provided, it is possible to look into the data in more detail.

Throughout this work, it has been pointed out how language can be described as an inventory of constructions. Constructions are the units on which speakers draw in order to communicate (Tomasello, 2003; Dąbrowska, 2004; Goldberg, 2006). According to usage-based models, children's language is concrete, in that children draw on constructions which are lexically-bound; i.e. built around specific lexical material. In other words, their constructions are either fully lexically-specific or partially schematic.

Assuming that such an account of children's early language is correct, tracing sentences uttered in the test corpus back to putative units attested in the main corpus makes it possible to identify the (putative) inventory on which a child draws. Since language is an inventory of constructions, a qualitative insight into the units on which the child may *potentially* rely is a qualitative insight into that particular child's grammar at a given point during linguistic development.

7.2.1. FULLY LEXICALLY-SPECIFIC STRINGS[45]

Fully Lexically-specific Strings are units which are fully phonologically specified and can be thought of as pre-packed, formulaic strings of various length. Of the 502 strings identified, 186 (37%) were **fixed-strings** which could not possibly present other orders (such as article-noun combinations) and only 6 (1%) were **fully-specific-packets** (i.e. units whose internal WO is only partially specified; table 7.7).

Table 7.7: Types of Fully Lexically-specific Strings.

Types of Fully Lexically-specific Strings	no.	%
fixed-strings whose internal order is the only possible grammatical sequence	186	37%
single words	156	31%
fixed-strings	147	29%
onomatopoeias	7	2%
packets	6	1%
total	**502**	**100%**

Fully Lexically-specific Strings could be of various length in terms of the numbers of words and morphemes and they were retrieved in production up to 13 times[46]. The longest *Fully Lexically-specific String* is the fixed-string reported in (50), which contains 14 morphemes and 7 words (morpheme counting is indicated in square brackets after each morpheme).

[45] Those units are the ones that have been used to successfully trace-back target sentences. Hence, units used to (partially) trace back, say, a lexical fail are not considered.

[46] I.e., the most common Fully Lexically-specific String was uttered 13 times by Roberto in the test corpus.

50. *C' era una volta una bella favolina.*
c(i)[1] *er[2]-a[3]* *un[4]-a[5]*
there(clitic.LOC) [1] be(IMPERF)[2]-3.SG[3] a[4]-F.SG[5]
volt[6]-a[7] *un[8]-a[9]* *bell[10]-a[11]*
time[6]-F.SG[7] a[8]-F.SG[9] nice[10]-F.SG[11]
favol[12]<in[13]>-a[14]
fairy.tale[12]<little(DIM)[13]>-F.SG[14]
"Once upon a time there was a little fairy tale."

7.2.2. SCHEMAS WITH SLOTS

The method identified 698 *Schemas-with-Slots*, 506 (74%) containing only one slot and 182 (26%) being two-slot schemas. Such schemas could be at the word (single-word schemas) or at the clause (multi-word schemas) level. In the latter case, they could be either *schematic-packets* or *fixed-schemas* (table 7.8).

Table 7.8: types of schema with slots.

Types of Schemas-with-Slot	no.	%
multi-word fixed-schemas	634	91%
single-word schemas	35	5%
multi-word schematic-packets	29	4%
total	698	100%

Overall, the method identified 870 slots, which can be grouped into two categories: **semantic** (no. 791; 91%) and **morphological** (no. 79; 9%) slots.

7.2.2.1. Semantic Slots

Semantic Slots represent semantic and/or functional generalisations across words or string of words that have a similar function in the schema's instantiations.

For instance, ***THING* slots** mainly represent generalisations across agents (*DOER*, *PUSHER*) and patients (*THING_DONE*, *THING_PUSHED*), but they may also represent generalisations across other functions (e.g. RECEIVER). They may be part of schemas which specify their gender and number (fig. 7.7a), rather than their size (fig. 7.7b), or part of schemas which specify no morphological feature. In this latter case, a *THING* slot can be specified as a singular (fig. 7.8a) or as a plural (fig. 7.8b) entity, or be instantiated by both plural and singular NPs[47] (fig. 7.8c).

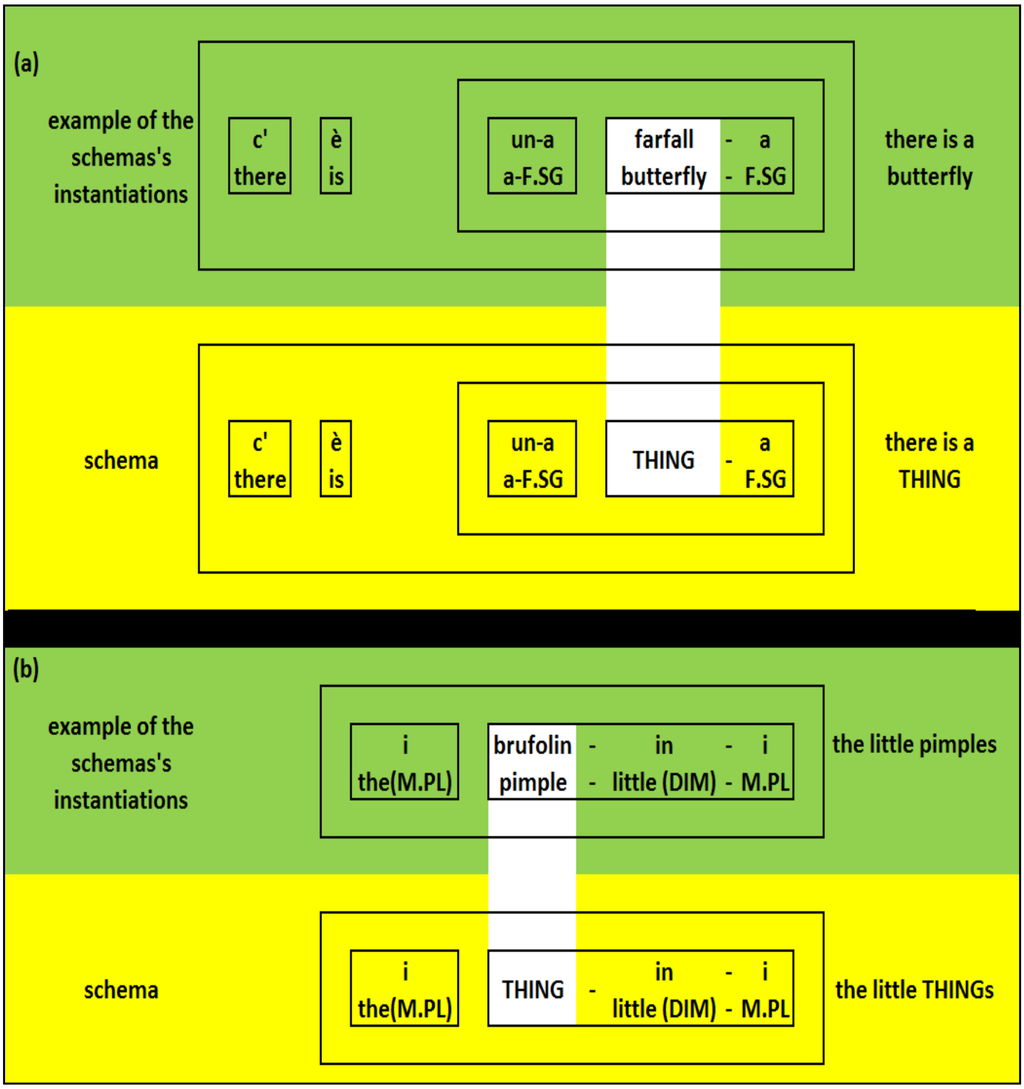


Figure 7.7: the THING slot (highlighted in white); 1 of 2. Schemas are highlighted in yellow and their instantiations are highlighted in green. Slot formation is highlighted in white.

47 Refer back to 6.4.5.1.

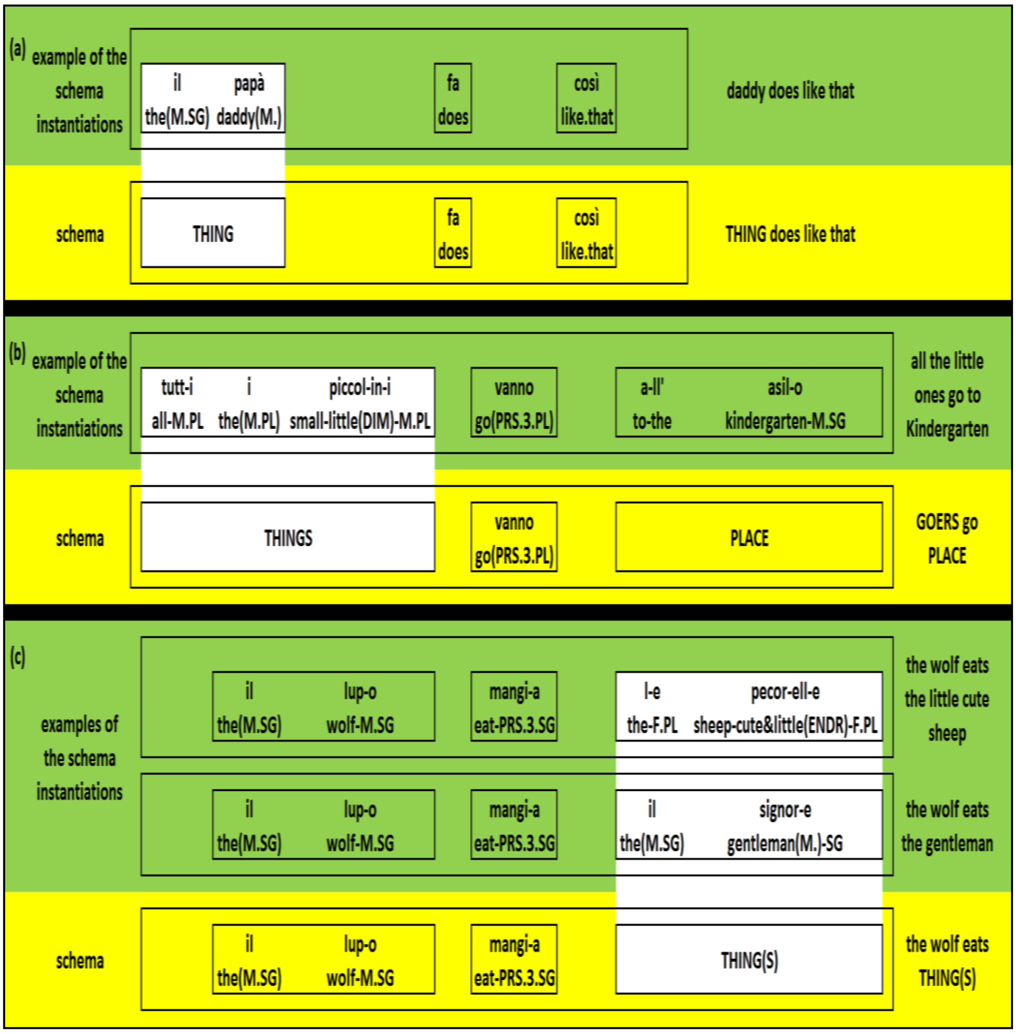


Figure 7.8: the THING slot (highlighted in white); 2 of 2. Refer to fig. 7.7 on how to read this figure.

Table 7.9 shows that the most attested semantic slots are *THING*, *PROCESS*, *PLACE* and *SENTENCE*, which together account for 96% of all slots.

Table 7.9: types of semantic slots.

Semantic slot type	no.	%
THING	426	54%
PROCESS	230	29%
PLACE	52	7%
SENTENCE	51	6%
QUALITY	14	2%
others	18	2%
total	**791**	**100%**

7.2.2.2. Morphological Slots

Morphological slots represent generalisations across bound morphemes that carry grammatical information such as gender, number, aspect, mood, tense and person and can be bound to roots of nouns and verbs.

INFLECTION slots are morphological slots which are bound to the roots of specific verbs. Such slots represent generalisations across various aspect-mood-tense-person-number inflections that can merge with verbs (fig. 7.9). About 68% of morphological slots represent generalisations across verbal inflections (as in fig. 7.9).

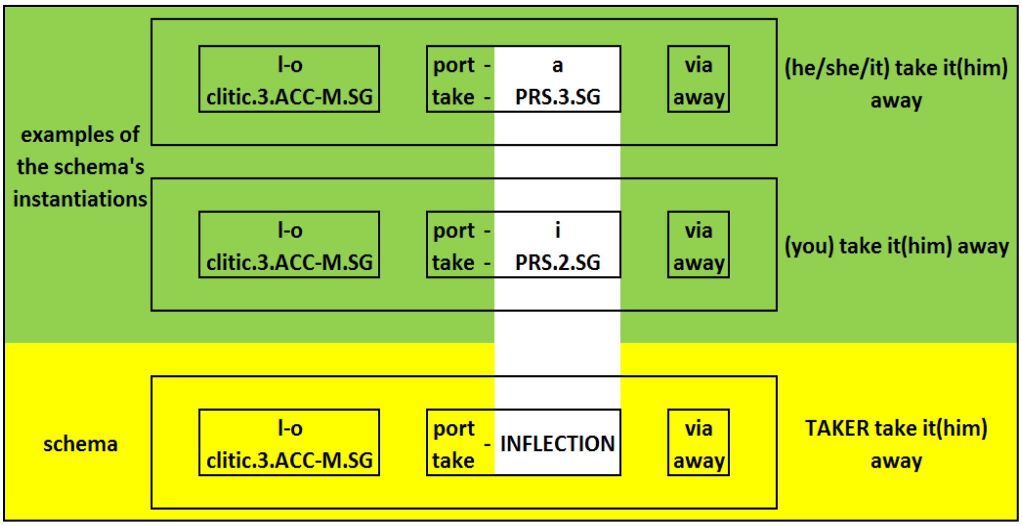


Figure 7.9: process(root)-INFLECTION slot (in white). Refer back to fig. 7.7 on how to read this figure.

The reader may find an accurate report of all slots yielded by the method, the most frequently retrieved ones, examples of the schemas of which they are part and some of their instantiations in Appendix_II (ch. 20.2.4.1-20.2.4.3).

7.3. CROSS-LINGUISTIC COMPARISON

Overall, the outcomes of this research fit well with the results of previous studies on the linguistic development of English-speaking children. Figure 7.10 compares Roberto's results with the results reported in Lieven et al. (2009) and Dąbrowska and Lieven (2005). Despite different test corpus sizes, frequency thresholds for available precedents, types of constructions investigated and languages enquired into, 79% to 93% of what children say can be accounted for in terms of lexically-specific constructions that they have already encountered during their previous linguistic experience. Overall, constructional fails represent 4.3% to 16.3% in Lieven et al. (2009, p. 492), up to 18% in Dąbrowska and Lieven (2005, pp. 451-455) and 13% of Roberto's tracebacks.

Looking at fig. 7.10 it is possible to note that about 40% of Roberto's tracebacks consist of multi-operation derivations, whereas such derivations represent 1% to 21% of Lieven et al.'s (2009) tracebacks (M= 12%) and 0 to 49% (M=15%) of the sentences analysed by Dąbrowska and Lieven (2005). Thus, Roberto's tracebacks present more multi-operation derivations than most of the tracebacks of English-speaking children.

Such a difference persists even when Roberto's sentences are analysed with Method_C (see Appendix_II, chapter 20.2), which is virtually identical to the one adopted by Lieven et al. (2009)[48]. Fig. 7.11 compares the results yielded by Method_C and the results of Lieven et al.'s (2009) study. The figure shows that Roberto's tracebacks still present a proportion of multi-operation derivations (39%) which is nearly twice as high as in the tracebacks of English-speaking two-year-olds (1%-21%). Hence, this cross-linguistic difference is not a by-product of the frequency threshold adopted to establish whether a string is classified as an available unit.

[48] However, Method_C allows precedent and target sentence to have different word orders (which was not allowed in Lieven et al., 2009).

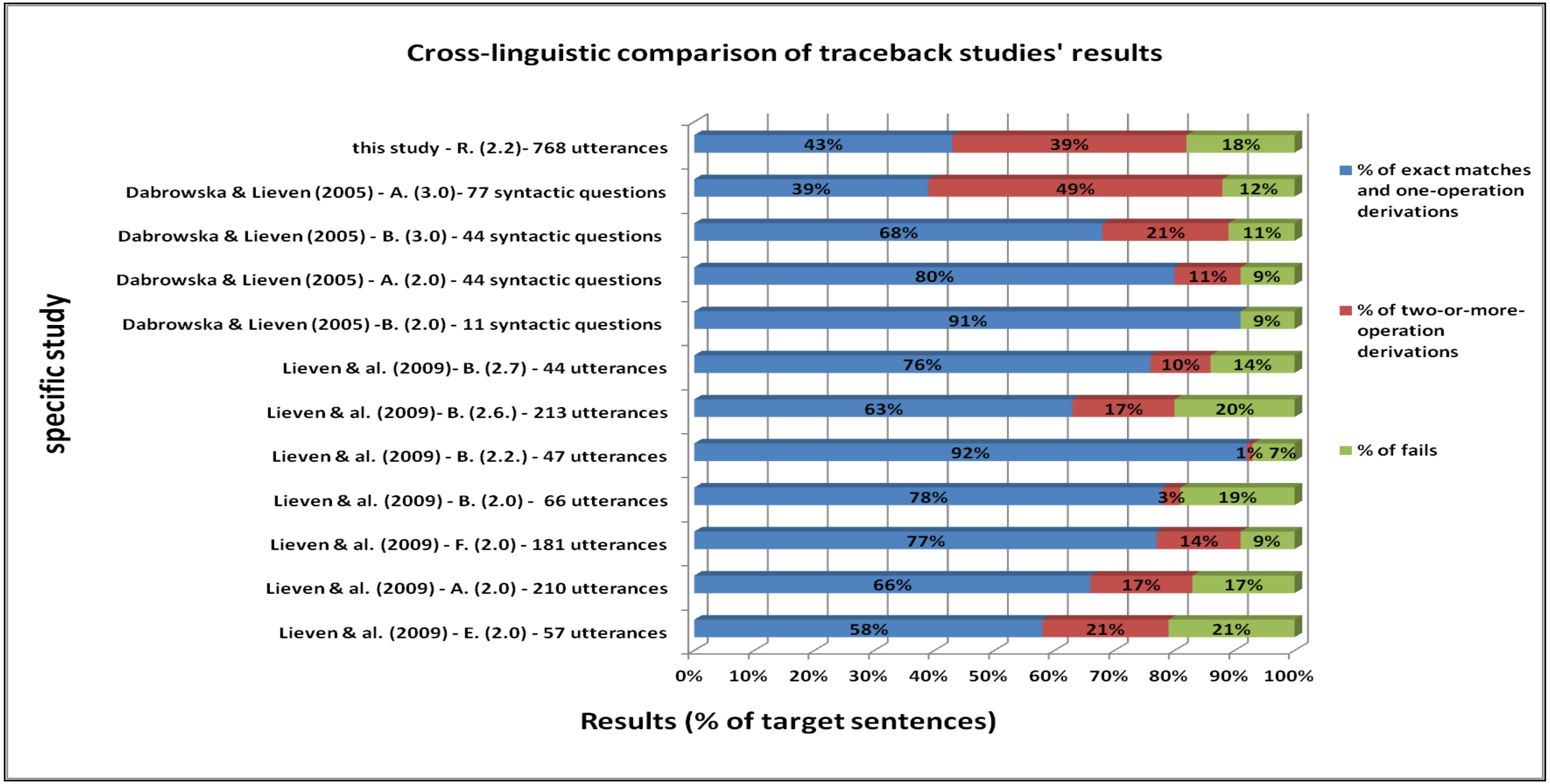


Figure 7.10: Comparing the results of various traceback studies. On the left of each bar it is indicated: the study to which the data belong (year of publication) – child's initial – (child's age) – type and number of constructions enquired.

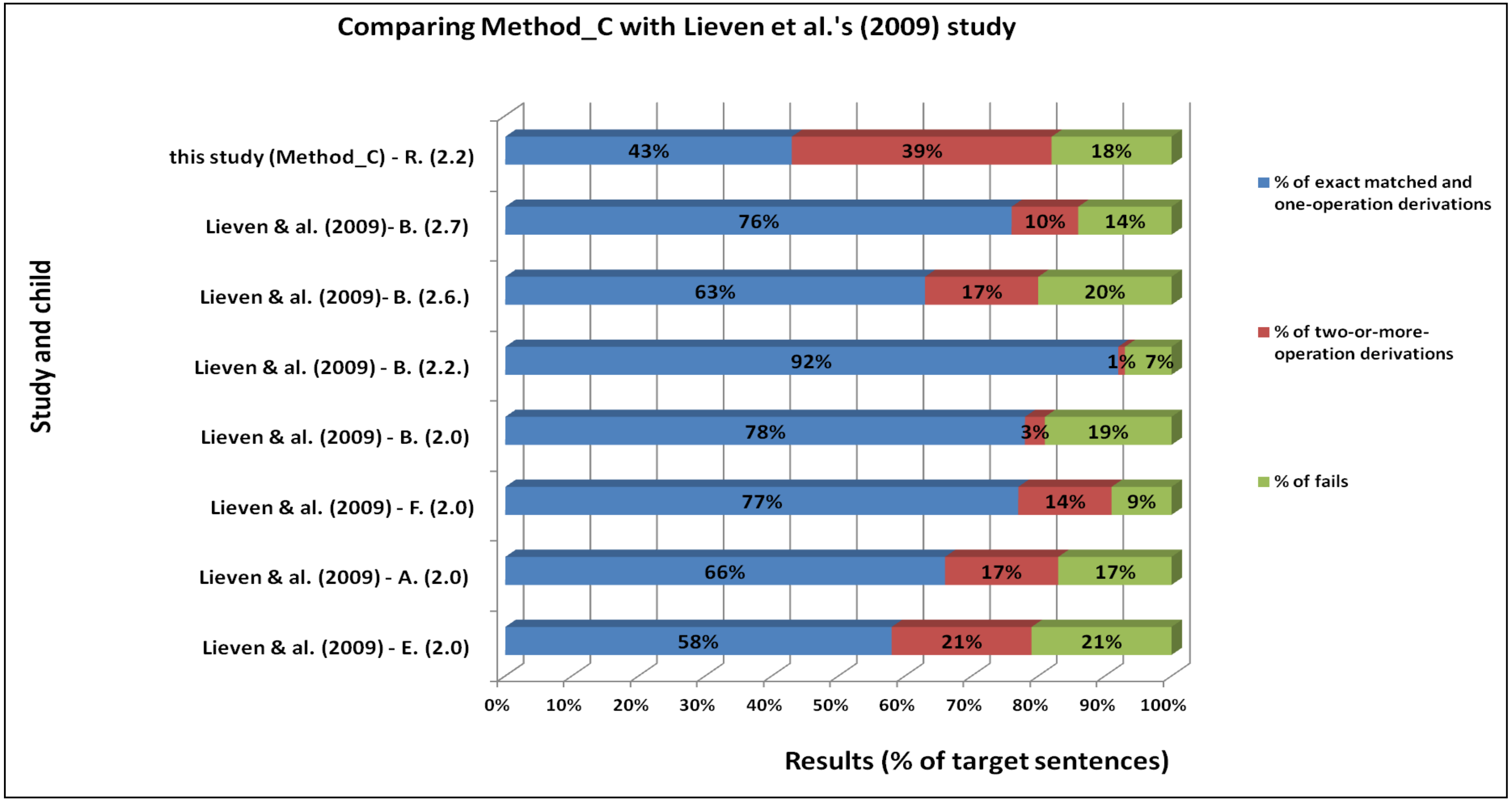


Figure 7.11: Comparing the results of Lieven et al. (2009) with Method_C. On the left of each bar it is indicated: the study to which the data belong (year of publication) – child's initial – (child's age).

It is however interesting to note that when Roberto is matched for MLU with English-speaking children, his tracebacks are fairly similar to the tracebacks of one English-speaking child. Both the results of Roberto's tracebacks and his MLU in words resemble those reported by Dąbrowska and Lieven (2005) for A. when she was three years old (fig. 7.12). Roberto's MLU in words is 3.14 (sd=1.96) in file A and 3.044 (sd= 2.04) in file B while the English-speaking girl's MLU is 3.48. The proportion of multi-operation derivations is 39% for Roberto and 49% for the girl. Overall then, Roberto's tracebacks are more similar to three-year-old A.'s tracebacks than to the ones of other English-speaking two-year-olds. Higher MLU means longer utterances, which in turn implies the ability to say more things (more words). Hence, it is not surprising that children whose MLU is higher display a higher proportion of multi-operation derivations. Indeed, Lieven et al. (2009, p. 496) found that as children's MLU increased so did the proportion of multi-operation derivations in their tracebacks. Not surprisingly then, the length in words of Roberto's target sentences and the number of operations needed to trace them back are strongly and positively related (r=.8, p(two-tailed)<.0001).

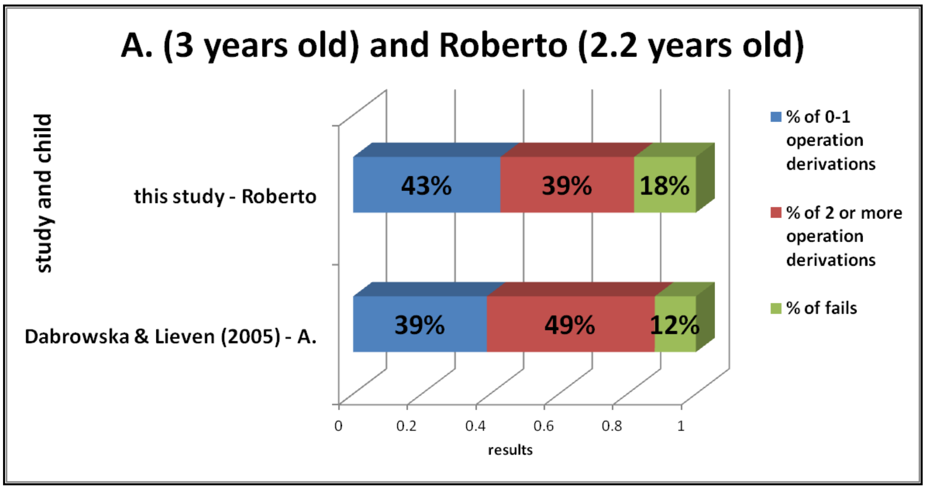


Figure 7.12: Comparing the results of A. (English-speaker, 3;00) and Roberto (Italian-speaker, 2;2).

However, such similar results have to be taken with caution, as Dąbrowska and Lieven (2005) enquired into a narrow set of constructions, namely, syntactic questions. Such constructions, as Lieven et al. (2009, p. 501) point out, are likely to have a high degree of lexical specificity, as they are bound to a narrow set of verbs (*be*, *do, shall*, etc.) and *wh* words. One may therefore argue that the interaction between the routinesed situations children experience and the specific constructions investigated by Dąbrowska and Lieven (2005) is likely to yield results that underestimate the degree of schematicity of children's language. Thus, it is indeed possible that if three-year-old A.'s entire production were considered, it might show a much higher proportion of multi-operation derivations than Roberto's target sentences.

Clearly, results from only one Italian-speaking child, let alone an Italian-speaking child whose vocabulary is impressively advanced, cannot be taken as conclusive findings that target sentences require more operations when the language enquired is Italian than when English is under analysis. Indeed, factors such as test corpus size and MLU might be variables that significantly contribute to yielding such results.

Unfortunately, a thorough investigation of cross-linguistic differences in children's tracebacks would take this discussion too far afield and is therefore a question for further research.

What is relevant for the purpose of the research question posed in chapter 5 is that results are cross-linguistically similar, both in terms of successful derivations and in terms of constructional fails.

7.4. SUMMARY OF RESULTS

Overall, 82% of Roberto's target sentences can be accounted for in terms of lexically-specific units extrapolated from the input (fig. 7.1, topmost bar; table 7.1). This leaves out a substantial proportion (18%) of target sentences that cannot be successfully derived (135 sentences). Fewer than 30% of unsuccessful derivations are lexical fails (48 sentences, 5% of all target sentences), whereas 67% of them are classified as **constructional fails** (90 sentences, 12% of all target sentences). A further 7 sentences (1% of all target sentences) contain both a constructional and a lexical fail (table 7.4). **Constructional fails** are those sentences that cannot be accounted for by applying superimposition and/or juxtaposition to precedents that are classified as available units. Constructional fails at the word level are called **morphological fails**, whereas constructional fails at the clause level are called **syntactic fails** (table 7.6).

Results are consistent with previous studies on English-speaking children (fig. 7.10), both in terms of successful derivations (79% to 93%) and in terms of constructional fails (0% to 21%).

The majority of sentences (61%) could be traced back by applying zero to two operations (fig. 7.1, topmost bar). Furthermore, 19% (144) of Roberto's utterances are strings attested at least once in the main corpus (sometimes with the same WO, sometimes with a different WO) and only 9% of utterances cannot be accounted for by what the child heard or said before (Method_D, fig. 7.1, bottom bar). Superimposition is overwhelmingly the operation through which target sentences seem to have been assembled (98%).

The method yielded 502 *Fully Lexically-specific Strings* and 698 *Schemas-with-Slots* to which target sentences were (successfully) traced back. *Schemas-with-Slots* could be at the word (5%) or the clause (95%) level. In the latter case, they could be either schematic-packets (4%), whose internal order was flexible, or fixed-schemas (91%), whose internal order was fixed. Schemas were allowed to have a maximum of two slots: 74% of them presented only one slot. Overall, 870 slots were identified, which could be classified as either semantic (91%) or morphological (9%).

8.

ANALYSIS

This chapter presents a more qualitative analysis of Roberto's target sentences, overwhelmingly focussing on constructional fails and how they can be interpreted within a UB framework. In order to help the reader to walk through this chapter, a guideline of how the analysis proceeds is provided below.

Section 8.1 focuses on two cases of morpho-syntactic overgeneralisation that are accounted for by using Method_A. If children rely on schemas and these are productive units, then over-generalisations are an expected outcome (see Dąbrowska, 2000).

Section 8.2 discusses Soft_Constructional_Fails at the word level (**morphological fails**). I shall argue that Roberto seems to master regular nominal inflections in a productive and mature way. Such an outcome is what a UBA to LA would predict, given Roberto's large vocabulary and the highly regular and frequent inflectional patterns of Italian.

Soft_Constructional_Fails at the clause level (**syntactic fails**) are then analysed as extensions vis-à-vis strings attested in the main corpus (8.3).

The analysis then moves on to Hard_Constructional_Fails (8.4), which present a higher proportion (34%) of ungrammatical sentences than both Soft_Constructional_Fails (15%) and successful derivations (2%). This is interpreted as evidence that children are creative learners. However, when such creativity is not supported by well-entrenched lexically-specific units, they are likely to struggle to produce adultlike, grammatical sentences. This claim is supported by:

a) the analysis of two sentences that can be interpreted as extensions vis-à-vis a prototype. Such extensions appear to be rooted in formal and functional, rather than syntactic and adultlike generalisations (8.4.3).

b) the analysis of grammatical agreement, which casts doubt on whether Roberto has acquired a "lexically-independent" competence of it. When various kinds of agreement attested in his target sentences cannot be explained in terms of lexically-specific schemas, the child actually produces more wrong (60%) than correct (40%) agreements (8.4.4).

The analysis ends with some observations about the nature of superimposition; in a highly inflected language such as Italian, it is reasonable to hypothesise that such an operation may also take the form of concatenations of partial overlaps (8.4.5).

8.1. THE EMERGENCE OF MORPHO-SYNTACTIC OVERGENERALISATIONS

By using Method_A it is possible to account for Roberto's morpho-syntactic overgeneralisations and creative usages of language. A case in point is represented by target sentences SB544 (51). Here, Roberto uses the intransitive *scendere* "to descend" or "to come/go down" in a transitive construction.

*51. * dai, Luca,* ***scendilo giù!***

da-i,	*Luca,*
give-PRS.2.SG,	Luca,
scend-i=l-o	***giù!***
descend-IMP.2.SG=clitic.3.ACC-M.SG	**down**!

"*Com'on, Luca, descend it/him down!"
"*Com'on, Luca, go it/him down!"
(week6.2014.02.18.B: line 3959)

Sentence (51) can be traced back to the schema in fig. 8.1 (yellow strip) and derived as in fig. 8.2.

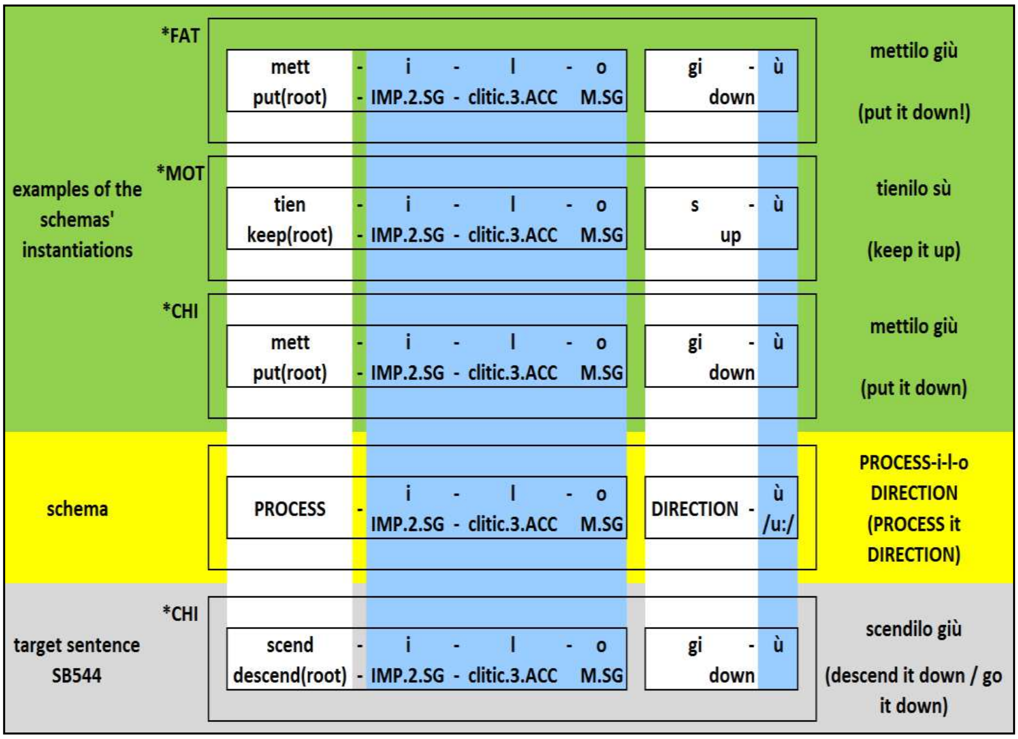


Figure 8.1: The schema (yellow) from which target sentence SB544 (grey) was derived and the schema's instantiations in the main corpus (green). Relationships of elaboration are in white; shared concrete material in blue.

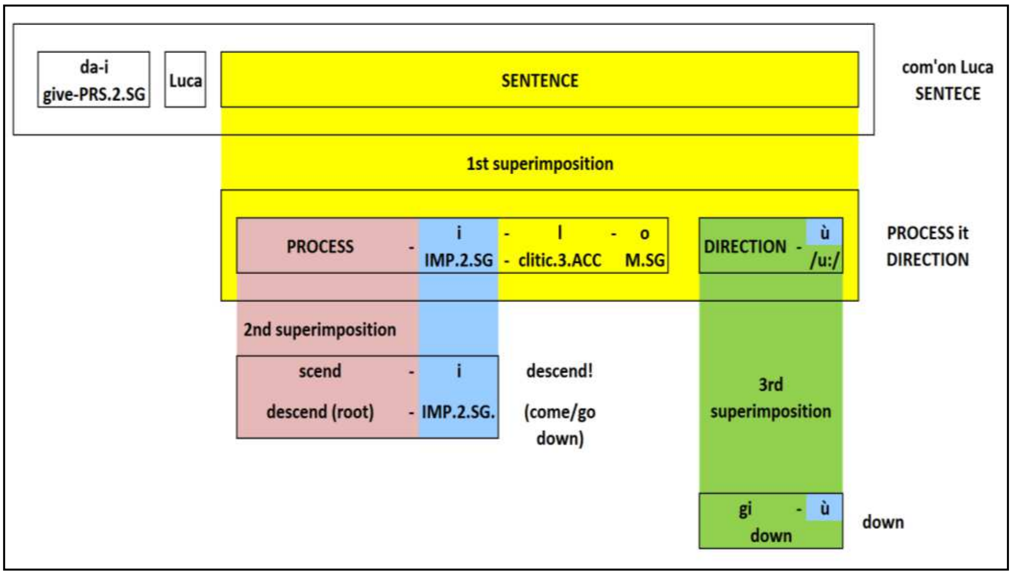


Figure 8.2: deriving target sentence SB544. Slot elaboration is in yellow, green and pink. Shared concrete material is in blue. Please note that the order of the superimpositions/operations in this and other figures is not meant to represent the exact order of assembly. The method makes no assumptions as to the order in which component units are assembled.

Interestingly, the schema in fig. 8.1 (yellow strip) is built around, amongst other elements, a case-marked (3.ACC) pronoun (*l*-) which merges with the verb in post-verbal position (given the imperative inflection). Several scholars (e.g. Tomasello, 2003; Akhtar, 1999) hypothesised that pronoun-based constructions

(*I'm-PROCESSing-it*) might constitute a foothold towards the development of constructional schemas.

Childers and Tomasello (2001) trained two-year-olds (2;4-2;10) with real transitive verbs, assigning participants to one of the following training conditions:

a) Children heard only NP-V-NP sequences (*Look! The cow's pulling the car. See? The cow's pulling the car*).
b) Children heard both NP-V-NP and PRON-V-PRON sequences (*Look! The cow's pulling the car. See? He's pulling it*)
c) No training.

During the test phase children were exposed to either intransitive or passive novel nonce verbs and transitive uses of those same verbs were then elicited. Children trained with pronouns were nearly twice as productive as children trained with NPs only ($F(1,40)=4.78$ $p<.05$). Furthermore, the former, but not the latter, outperformed the no training group ($p<.05$). The implication is that children in group B could strengthen their pronoun-based schemas and use them to produce adultlike transitive sentences with novel verbs.

Similarly, children who participated in other experimental studies (e.g. Akhtar & Tomasello, 1997; Akhtar, 1999) managed to overcome ill-formed input by relying on pronoun-based constructions.

Thus, experimental evidence appears to be consistent in indicating that pronoun-based schemas play a crucial role in children's early creative (productive) utterances.

Recall section 3.9.1; children were able to draw functional connections between the truck in picture B and the car in picture A. Tomasello (2006b) reports that children were even more successful with the task whenever the agent/tow-er was identical (a car) in both pictures. Tomasello (2006b) suggests that functional analogy is easier when some elements of the structures being analogised (picture A and picture B) not only have the same function (tow-er), but they also are perceptually identical (they are both cars). The case-marked nature of pronouns, besides making them local cues (hence easier to acquire than more global cues), provides elements of phonological and functional identity that are likely to facilitate analogy since specific forms (*I*, *she* versus *me*, *her*) are constantly associated with a specific function (*agent* versus *patient*) and a specific distributional pattern (pre and post-verbal position). Furthermore, pronouns are very frequent in CDS (Tomasello, 2003). Thus, the interaction of clear form-function mapping and frequency is likely to create the conditions for pronoun-based schemas to become entrenched (and productive) constructions very early on in development.

This is likely to be even truer in Italian, in which the post-verbal position is not as reliable a cue to patient role as in English (see Bates & MacWhinney, 1987). In Italian, about 30% of *V-NML* sequences instantiate *VS* patterns (Dell'Orletta, Lenci, Montemagni & Pirrelli 2005; Bates, 1976), whereas accusative clitics constantly map onto patient role.

Hence, given that imperatives are fairly frequent in CDS (Abbot-Smith & Serratrice, 2013; Tomasello, 2003), schemas such as *PROCESS(IMP)=l-o* "PROCESS(IMP)=clitic.3.ACC-M.SG" are excellent candidates to become entrenched (and reliable) formulas.

Overall, it may then be the case that in the same way as English-speaking children rely on pronoun-based schemas, Roberto draws on the schema in fig. 8.1 (yellow strip) to use the intransitive *scendere* "descend" transitively[49]. Thus, a UBA can accommodate and account for sentence (51) in terms of lexically-bound generalisations.

However, this is not Roberto's whole story. In the **main corpus** another case of overgeneralisation (52) is found.

52. *CHI: **hai (...) caduto il cagnolino dal trattore .*

hai	*cad-u-to*	*il*
have(PRS.2.SG)	fall-TV(conj.II)-PTCP	the(M.SG)
cagnol-in-o	*da=l*	*trattor-e*
dog-little(DIM)-M.SG	from=the(M.SG)	tractor(M.)-SG

"You have fallen the little dog from the tractor."
(week4.2014.01.28: line 2114)

[49] Note though, that in some varieties of Southern Italian, *scendere* can be used transitively.

Clearly, this sentence is the best candidate for arguing that Roberto is relying on a fully schematic (*AGENT-PROCESS-PATIENT*) representation. He knows that, in order to express a transitive action, he has to utter the sequence VO. Since he has not learnt which kinds of arguments *cadere* "to fall" can take (i.e. that it is an intransitive only verb), the sentence in (52) is uttered.

There is also another interpretation, though; one that does not involve positing a fully-entrenched schematic representation.

Both (51) and (52) map onto scenes in which X (*2.SG*) causes Y (*the little dog*; *it*) to move Z (*from the tractor*; *down*). That is, they both instantiate the **caused-motion-construction (cmc)** (see Goldberg, 1999).

The reader may refer back to section 3.10.2 and Goldberg's prototype-based model of generalisation. Semantically similar verbs appear in the same Argument-Structure-Construction. Each Argument-Structure-Construction is dominated by a light verb (e.g. *give* dominates the ditransitive), which can be thought of as the prototype of the complex category representing knowledge of that specific constructional pattern.

A working hypothesis, discussed in section 3.10.2, is that new patterns instantiating a particular Argument-Structure-Construction (*email her a draft*) are initially apprehended as extensions vis-à-vis a lexically-specific pattern (*give-RECIPIENT-THING_GIVEN*) perceived as the prototype of the construction (*TRANSFER-NML1-NML2*). In doing so, a superordinate schema instantiated by both new expression and prototype is fleetingly abstracted (fig. 8.3; see also section 2.5.2). Initially, such a superordinate schema is only weakly entrenched and can be evoked (or accessed) only via the activation of more specific units. As more and more lexically-specific patterns are apprehended (learnt) as extensions vis-à-vis the prototype, this more schematic unit (A' in fig. 8.3) gradually acquires representational strength.

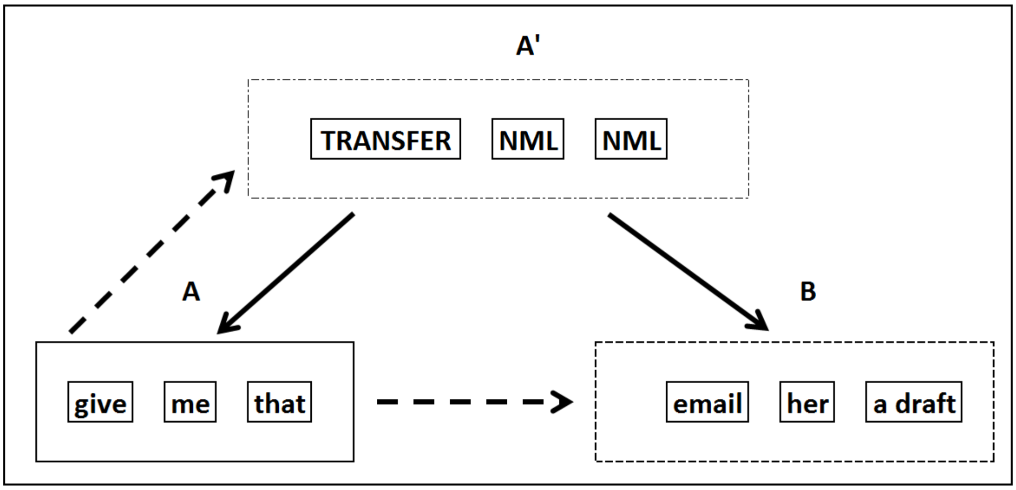


Figure 8.3: Apprehending an instance of the ditransitive construction (B) as an extension vis-à-vis the construction prototype (A). Dashed arrows indicate relationships of extension and solid arrows indicate relationships of elaboration. Thickness of boxes indicates degree of entrenchment.

Goldberg et al. (2004, p. 303) point out that their hypothesis concerns how learners might draw analogies across instances of similar patterns; it is not a model of productivity (how and when speakers use an Argument-Structure-Construction in production). Nevertheless, the two issues are strictly related; having some kind of representation of a given construction is essential to start using it productively.

Returning to Roberto's sentences, it is worth observing that

a) Both (51) and (52) are instantiations of the caused-motion-construction (cmc).

b) According to Goldberg (1999), *put* represents the prototype of the cmc.

c) Target sentence SB544 (51) could be interpreted as an extension vis-à-vis a string attested three times in the main corpus (fig. 8.4). The main corpus string (A in fig. 8.4) differs from SB544 (B in fig 8.4) only in that it uses a different verb, namely *mettere* "to put" (fig. 8.4). Thus, SB544 can be thought of as an extension vis-à-vis a prototypical (lexically-specific) pattern.

d) In tracing Roberto's sentences back to their putative component units, four lexically-specific schemas which were built around *mettere* "to put" *and* which instantiated the cmc were identified (see Appendix_II, ch. 20.3.1).

e) Tomasello (2003) hypothesises that abstract constructions develop from structural and semantic alignment across specific units. This form-function correspondence seems to be present across the sentences under analysis (fig. 8.5; the agent of the action is indicated on the verb, as it is expressed morphologically).

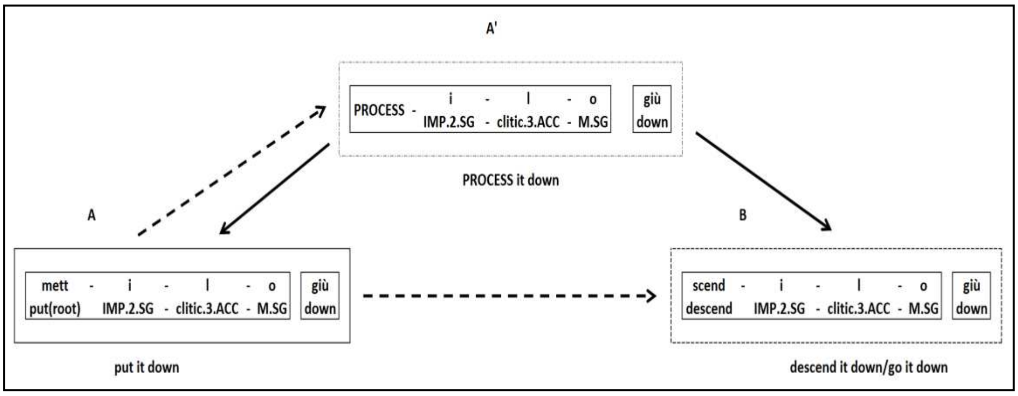


Figure 8.4: Producing target sentence SB544 (B) as an extension vis-à-vis a prototypical instantiation of the cmc (A). See figure 8.3 on how to interpret lines, arrows and boxes.

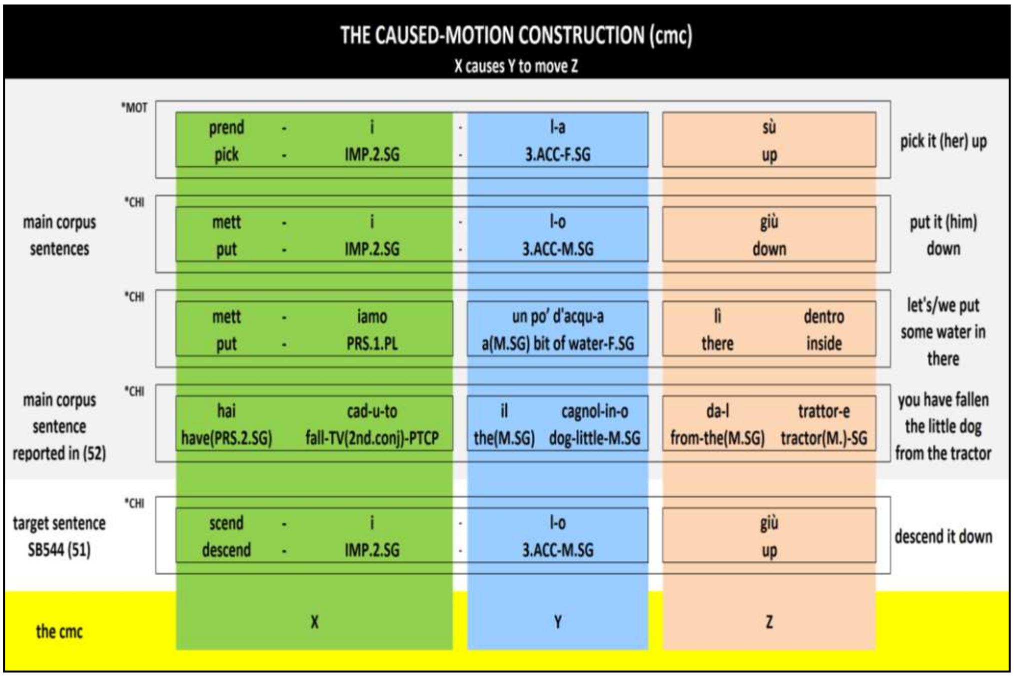


Figure 8.5: the caused motion construction and (some of) its instantiations in the corpus collected.

The hypothesis put forward here is that (52) can be accounted for by the interaction of two factors:

a) Roberto is developing the caused-motion-construction in a piecemeal fashion out of lexically-specific schemas mainly built around the verb *mettere* “to put”.

b) The input *does* provide lexically-specific units out of which the sentence could have been built.

As for point (b), fig. 8.6 reports the lexically-bound units which could be regarded as the putative precedents of (52) attested twice or more in the recordings preceding it. At that point, only fifteen hours and twenty-nine minutes had been recorded

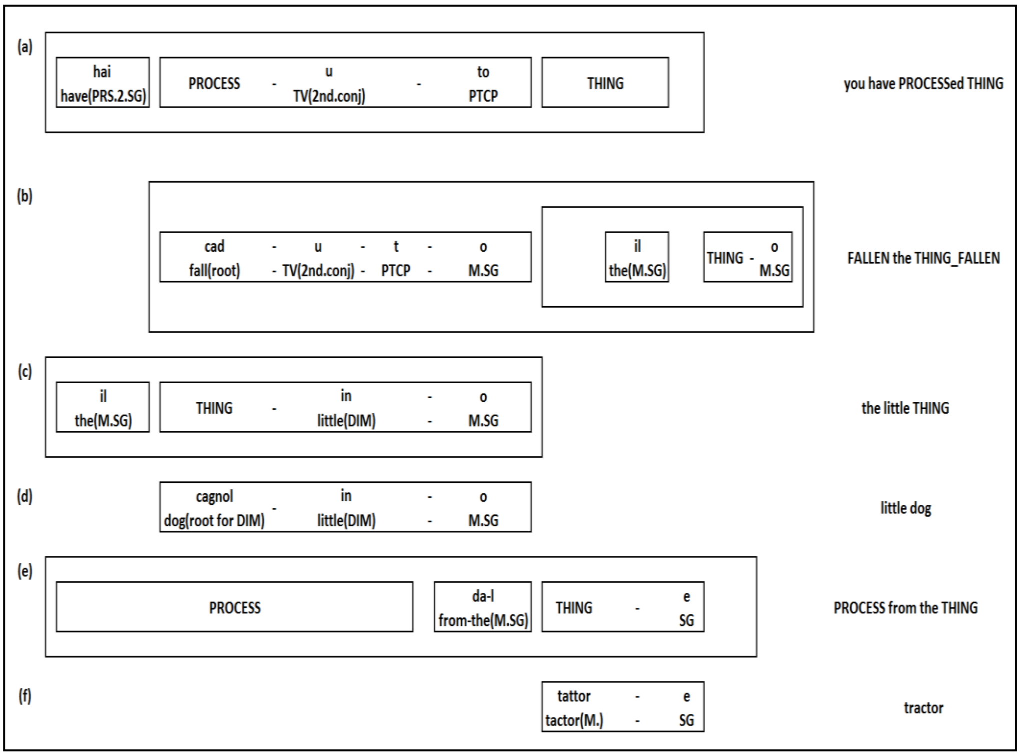


Figure 8.6: the units to which the sentence in (52) is traced back.

One might wonder what the difference is between the schema in fig. 8.6a and a more general transitive schema (*AGENT-PROCESS-PATIENT*), as both verb and object are left unspecified in the former as well as in the latter. I would argue that the difference is quite substantial. Firstly, the auxiliary in fig. 8.6a specifies the agent, which can only be 2.SG. Hence, the schema maps onto an event in which the addressee (2.SG) acts upon someone or something. Moreover, the fixed past participle suffix specifies that the event is in the past and also narrows the type of verbs that can fill the slot down to those whose past participle presents the TV *–u-*. This kind of schema mirrors the example made by Childers and Tomasello (2001, p. 740) *I-am-PROCESSing-X*.

Fig. 8.7 graphically represents the derivation of (52). The unit *il cagnolino* "the little dog" is attested only once in the first sixteen hours of recording. Hence, it has to be assembled (8.7a) in order to fill the slot of *fallen the THING_FALLEN* (8.7b). The resulting unit fills both slots of fig. 8.6a (see 8.7c). The PP is then added by means of two superimpositions (8.7d-e).

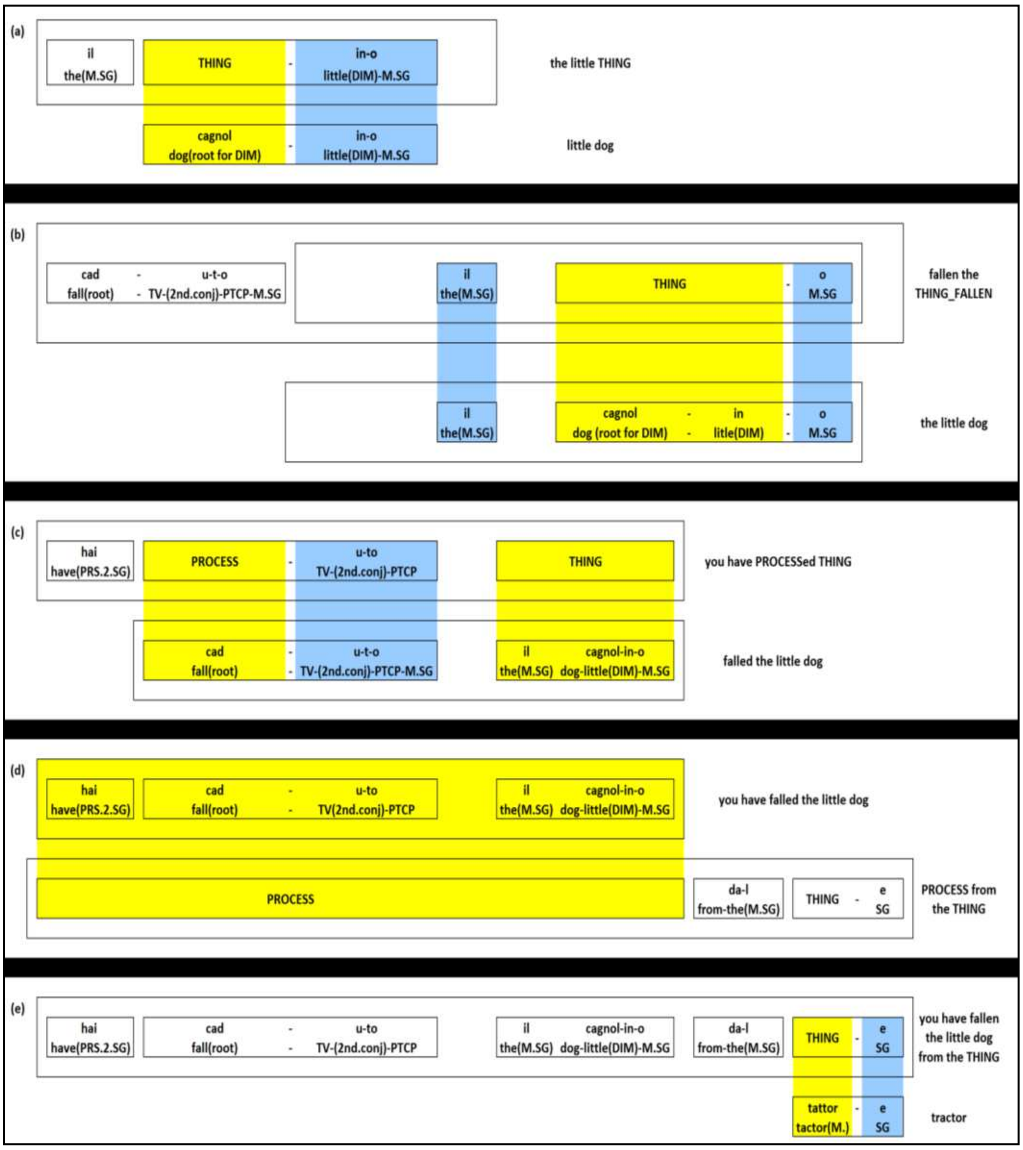


Figure 8.7: deriving the sentence in (52). Slot elaboration is highlighted in yellow and shared lexical material is highlighted in blue.

As for the two superimpositions in fig. 8.7c, an attempt is now made to hypothesise the underlying processes behind them. Fig. 8.6b (*fallen the THING_ FALLEN*) is a unit attested in the recordings preceding (52), with and without the grounding predicate (which should be the auxiliary *essere* “to be”). All throughout this study it has been repeatedly observed that a construction is a form-meaning pairing that maps onto a particular “scene”. Fig. 8.6b maps onto the end state of a *THING*’s downwards movement. The semantics of *cadere* “to fall" implies an entity that *undergoes* a movement; there is hardly anything voluntary in the action of falling (*fall* is unaccusative). On the contrary, other intransitive verbs of movement, such as *run*, can also imply a voluntary action in which an entity *undertakes* a movement (*run* can be ergative). That is, *falling* is something that happens to something or someone. *Running* is something someone does. Within the caused-motion-construction, the patient undergoes an involuntary movement. Hence, there seems to be a semantic correspondence between the DO of the caused-motion-construction and the *THING_FALLEN* slot of fig 8.6b; they both *undergo* a movement.

Such a connection might be strengthened by the fact that, in Italian, both intransitive subjects and transitive objects mostly take post-verbal position (Lorusso et al., 2005; Bates, 1976; Antinucci & Cinque, 1977). Hence, there is a distributional pattern, both in the ambient language and in the lexically-specific schemas yielded by the method, which reinforces a semantic link between the caused-motion-construction and the action of falling.

Roberto witnesses a life event where his interlocutor makes a little dog fall from a toy-tractor. It is possible that three units are activated: the (still weakly entrenched) caused-motion-construction, *caduto il THING-o* (fig. 8.6b) and *hai PROCESSuto THING* (fig. 8.6a). There are three important convergences that make (52) possible:

a) Fig. 8.6a matches the caused-motion-construction in that someone (the addressee) does something to something/someone. This something is causing the entity acted upon to move. Hence, the fully schematic caused-motion-construction might be accessed (retrieved) via a lexically-specific unit that (partially) instantiates it.

b) *THING_FALLEN* in fig. 8.6b and the object of the caused-motion-construction *undergo* a movement and both take post-verbal position. Again, the caused-motion-construction might be accessed (retrieved) via a lexically-specific unit that (partially) instantiates it.

c) The lexically-based units in fig. 8.6a and 8.6b match formally and semantically:

> The root of *cadere* (fig. 8.6b) matches the profile specified by the *PROCESS* slot of fig. 8.6a and the two units share the morphological ending (past participle) *–uto*, which maps onto a finished action ([+TELIC]). *THING _FALLEN* (fig. 8.6b) and *THING* (fig. 8.6a) are both *THINGS* in Langacker's (1987, 2000, 2008) terms and hence have matching profiles that can be elaborated by *il cagnolino* "the little dog". Both units present a VN sequence and both *THINGs* **undergo** some kind of action.

The fact that (52) can still be explained in terms of lexically-specific units (fig. 8.7) suggests that Roberto's linguistic representation is still very concrete. At the same time however, both overgeneralisations (51 and 52) are instantiations of the caused-motion-construction. This could be interpreted as evidence that Roberto is starting to draw analogies between lexically-specific schemas on the basis of similar form-function mapping (i.e., as evidence that he is developing a competence that is gradually becoming less and less lexically-bound).

One may argue that the fact that the above analysis hypothesised that the caused-motion-construction might have been retrieved (or accessed) via more specific patterns is itself evidence that Roberto's linguistic representation is somehow more abstract than the lexically-specific units yielded by the method would assume. I would argue though that it does not represent conclusive evidence.

Firstly, even if Roberto had developed a non-lexically-bound representation of the caused-motion-construction, such a competence per se would not be enough evidence that all constructions of his inventory have the same degree of abstraction. Constructions develop independently and different constructions may have different degrees of abstraction. Hence, the fact that Roberto might have developed the caused-motion-construction does not necessarily mean that other constructions (passives, questions) are not better accounted for as a set of related, yet independent, lexically-bound schemas. For example, there is evidence that English-speaking children develop schematic representation of the transitive construction earlier than the passive construction, often interpreting the latter as SVO patterns (Bever, 1970a-b; mentioned in Slobin and Bever, 1982).

Secondly, the very fact that all cases of syntactic overgeneralisation can be explained in terms of lexically-specific units appears to suggest that Roberto still relies on concrete constructions, rather than on fully-schematic templates.

Ultimately though, two cases of overgeneralisations uttered by the same child do not represent sufficient data from which one could draw certain conclusions.

The next section attempts to shed light on those sentences that are classified as constructional fails and what kind of insight into Roberto's linguistic competence they can provide.

8.2. MORPHOLOGICAL FAILS

Morphological fails – which represent about 11% of all fails (table 7.6) – constitute special cases of Constructional Fails: they are Constructional Fails at the word level. When a target sentence contains a word that is attested in the main corpus in one (morphological) form only and the particular form used by Roberto does not match exactly, this constitutes a morphological fail (refer back to 7.1.3.2 and table 7.6). Table 8.1 shows that two sentences contain both a morphological and a syntactic fail. Since in both cases the morphological fail appeared in combination with a Hard_Constructional_Fail, those two sentences were classified as Hard_Constructional_Fails.

Table 8.1: Distribution of morphological fails.

fail type	no.	% of 15 morphological fails	% of 135 fails	% of 768 target sentences
morphological fails	11	74%	8%	1.5%
both syntactic and morphological fails	2	13%	2%	0.3%
both morphological and lexical fails	2	13%	1%	0.3%
total	**15**	**1**	**0.11**	**2.10%**

Overall, the traceback analysis appears to indicate that Roberto has developed the ability to inflect (and therefore manipulate) nouns for gender and number, irrespective of whether he encountered evidence that those nouns can be inflected. Although I shall return to the early emergence of morphological productivity in more detail in the final discussion (chapter 16.2), it is worth pointing out a few factors that may account for such precociousness.

Firstly, there are language-specific features that may facilitate form-function mapping and hence schematisation. Many scholars (Akhtar & Tomasello, 1997; Lieven & Tomasello, 2008; Lieven & Brandt, 2011) note that the co-occurrence of high type and token frequency facilitates schema extraction at both the word and clause level. The Italian nominal inflectional system is highly regular (type frequency) and frequent (token frequency). Regular vowels map onto specific gender-number information (*-a* "F.SG"; *-e* "F.PL"; *-o* "M.SG" and *–i* "M.PL") and apply, apart from few exceptions and irregularities, across nouns, adjectives, past participles, articles and 3.ACC clitic pronouns. Indeed, all morphological fails[50] are regular items to which the four-vowel paradigm applies. This appears to suggest that the co-occurrence of regularity and frequency may have played a role in Roberto's early productivity. Indeed, Devescovi et al. (2005) found that, when English and Italian children are matched for vocabulary and age, Italian children display a more advanced morphological behaviour, which mirrors the greater weight that morphology has in their ambient language and, possibly, its clearer form-function correspondences.

Secondly, Roberto is a twenty-six-month-old precocious learner, whose vocabulary score is above the 75th percentile reported for thirty-month-olds. Marchman and Bates (1994) provided solid evidence that morphological productivity is strongly and positively related to vocabulary growth (refer back to 3.9.2).

Finally, Roberto's behaviour appears to be guided by the lexically-specific constructions on which he supposedly relies. In nine out of the fifteen sentences that are classified as morphological fails, the schema the child uses **does** account for the particular inflection taken by the noun (as opposed to the one attested in the main corpus). For example, (53) reports the relevant part of target sentence SA079, in which the child uses the word *cerv-i* "deer-M.PL", even though only the singular *cerv-o*; "deer-M.SG" is attested in the main corpus. All the child has to do is to impose the M.PL morphological marker of the schema over the M.SG singular ending of the word *cervo* "deer", so that the former overrides the latter (fig. 8.8, in the circle and in red)

[50]Fourteen nouns and one pronoun.

53. *I cervi vanno a nanna*

i	*cerv-i*	*vanno*	*a*	*nann-a*
the(M.PL)	deer-M.PL	go(PRS.3.PL)	to	beddy.bye-F.SG

"The deer go to sleep."
(week6.2014.02.18.A: line 719-720)

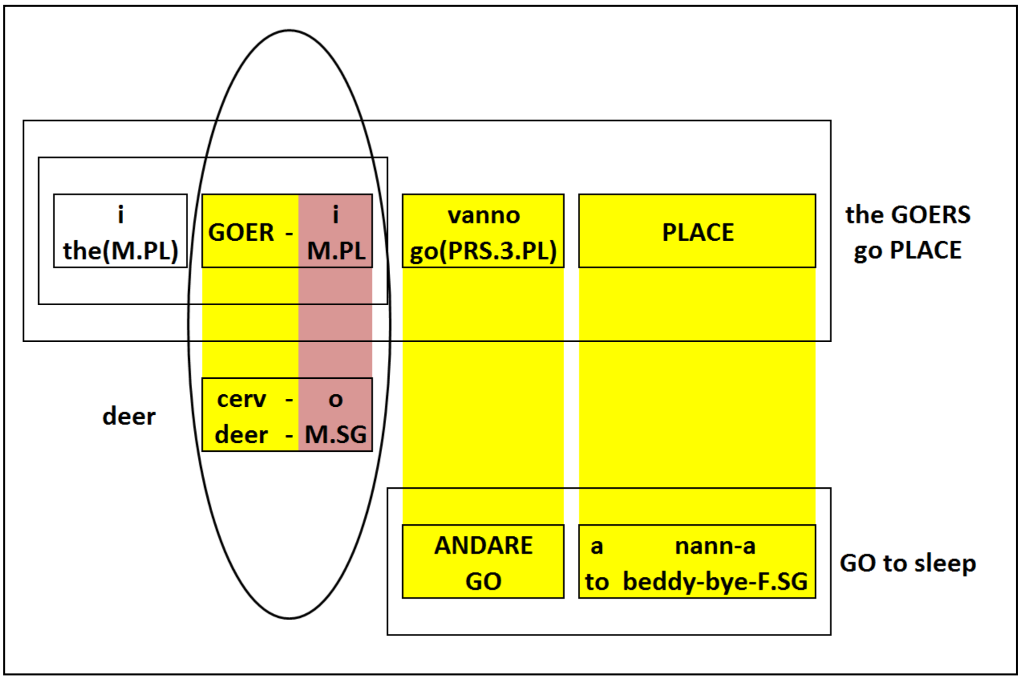


Figure 8.8: failing to derive part of target sentence SA079 (53). Slot elaboration is in yellow. Contrasting morphological specifications are highlighted in red.

Overall then, morphological fails, rather than representing problematic sentences, are in line with what a UBA would predict, given Roberto's large vocabulary, the morphologically rich linguistic environment surrounding him and the fact that most schemas indicate the specific gender-number markers to be used.

8.3. SOFT_CONSTRUCTIONAL_FAILS AT THE CLAUSE LEVEL: Syntactic fails

When a target sentence that **cannot** be accounted for by using Method_A **can** nonetheless be accounted for by using Method_C and/or Method_D, this is classified as a Soft_Constructional_Fail. More plainly, a Soft_Constructional_Fail is a target sentence that can be accounted for by relaxing the threshold for a precedent to be considered as an available component unit to one occurrence in the main corpus, whether this has been produced by the child (Method_C and Method_D; 35% of constructional fails, thirty-four sentences) or by any other speaker (Method_D; 13% of constructional fails, thirteen sentences).

About half (48%, forty-seven sentences) of constructional fails are classified as soft. Thirteen sentences are morphological fails (word-level fails) and thirty-four can be labelled as **syntactic fails** (clause-level fails). As previously discussed (6.4.2.1), the sampling regime adopted here (five/six hours per week) captures about 5% to 10% of a child's linguistic experience. Tomasello and Stahl (2004) showed that when a construction is as frequent as fourteen times a week, such a sampling regime has a 60% chance of capturing *ONE* instance of such a construction. It is therefore possible that these fails are by-products of the sampling regime and that Roberto actually heard those constructions with enough token and type frequency to infer generalisations from them. It is also possible though, that the strings caught on tape are the only occurrences of those constructions encountered by Roberto. In the following discussion the latter possibility is assumed to be the case and Soft_Constructional_Fails are interpreted as extensions vis-à-vis strings attested in the main corpus.

8.3.1. EXTENSION, SCHEMATISATION AND LINGUISTIC PROCESSING

Language competence is about mastering an inventory of symbolic units (form-meaning pairings), varying in schematicity and complexity. Speakers rely on those units to voice their communicative and interactional needs and wills (refer back to chapter 2).

Status of units is a function of **entrenchment**, which in turn can be thought of as being a function of **frequency** and **cognitive salience** (Langacker, 2000, 2008). Status of unit is then a matter of degree: the more a pattern is evoked, the more entrenched it gets; the more entrenched it gets, the easier it is to retrieve and use again. Langacker (2008, 2010) notes that the representation of schemas and units as clearly delimited boxes adopted thus far is misleading, as it imposes on them a discreteness they do not have. Language production and linguistic knowledge, as well as the relationships units have with each other (categorisation, symbolisation and composition), are ultimately **cognitive processes**.

The basis of categorisation is that a new experience (a new target structure (TS) “B”; *I gorp you*) is apprehended (sanctioned) on the basis of **previous experience** (a sanctioning structure (SS) “A”), because the TS is perceived either as an instantiation of the SS (*AGENT-PROCESS-PATIENT* ⟶ *I gorp you*) or as an extension of it (*DOER-do-THING* ⇢ *I gorp you)*.

For that categorising process to take place, such **previous experience** must be remembered. That is, the SS must leave a trace in speakers’ minds by which it can be remembered (Langacker, 2000, 2010). Once a trace has been left, such previous experience influences (sanctions) the apprehension of new experience (new target sentences). Since sanctioning a new structure (the unknown) on the basis of previous experience (what is known) is a **processing activity,** the more such an activity is undertaken, the easier it is to engage in again. Hence, the more entrenched a unit is, the easier it is to activate and use to categorise new expressions.

When the relationship between A and B is fairly straightforward, A is fully recognised in B (i.e. B elaborates A). This is the relationship of **full-schematicity** (or **instantiation** or **elaboration**) between a schema and its instantiation (*AGENT-PROCESS-PATIENT* ⟶ *I gorp you*). When, in order to recognise A (SS) in B (TS), one has to adjust and suppress some of A’s specifications, this is a case of **extension.** In order to apprehend the TS [[CIRCULAR ARENA]/[rɪŋ]] as a case of the SS [[CIRCULAR OBJECT]/[rɪŋ]], one has to override the OBJECT specification of the latter. In doing so, a superordinate structure A’ that bears a relationship of full-schematicity with both A and B is created, even if only momentarily (fig. 8.9).

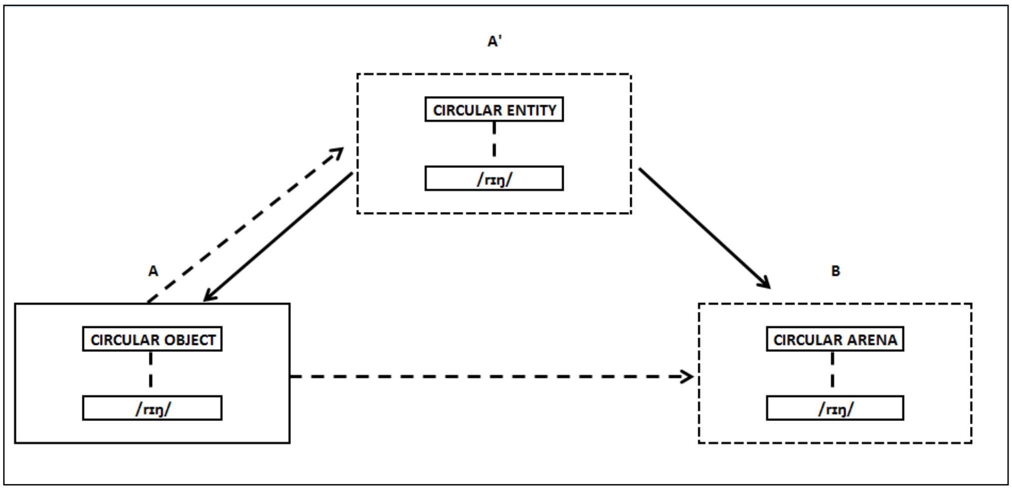


Figure 8.9 categorising B as an extension from A (from Langacker, 2000, 2008). See fig. 8.3 on how to read this figure.

In an attempt to enquire into similar approaches to language (CG, Emergent Grammar, Exemplar Theory and Analogy-based accounts), Langacker (2010) discusses the similarities between analogy-based accounts of grammatical generalisations and the processes of extension and schematisation in CG. Analogy-based accounts of grammatical generalisations posit that speakers create (and interpret) new sentences by drawing connections (analogies) across similar phenomena/sentences. Hence, a speaker who hears *I gorp you* can interpret this sentence by drawing analogies between the new sentence and a previously heard one, such as *I kick* you. Such a process correponds to a categorisation through extension (*I kick you* --► *I gorp you*).

However, analogy (and therefore extension) is possible because some similarities between the two sentences are perceived. Apprehending such similarities is what schema formation is about.

What then, is the difference between instantiation and extension?

In short: the entrenchment of the superordinate unit A'.

The apprehension of B as a case of A through extension implies the transient creation of the superordinate structure A', which is schematic with respect to both A and B. Like the weak transitive schema previously discussed, A' is evoked (accessed) only via A (i.e. an entrenched unit) in which it is immanent. Conversely, a schema with status of unit is presumably entrenched enough to be activated independent of its instantiations in order to sanction a new structure:

(A’ $\rightarrow$ B).

If (A $\dashrightarrow$ B) recurs enough times, it becomes entrenched and assumes status of unit: [A $\dashrightarrow$ B]. If, in the process, the superordinate unit (A’) acquires **cognitive salience**, it can then become an entrenched schematic unit [A’] that can be evoked independent of [A] to sanction new expressions[51]: (A’ $\rightarrow$ C).

8.3.2. SOFT_CONSTRUCTIONAL_FAILS AS CASES OF EXTENSION

Sentences that are classified as Soft_Constructional_Fails can be thought of as extensions vis-à-vis strings attested in the main corpus one time or more.

For instance, target sentence SB249 (fig. 8.10c and 8.11e)[52] can be derived from the schema in fig. 8.10b, whose only instantiation in the main corpus is fig. 8.10a. The derivation of target sentence SB249 under both Method_C and Method_D is represented in fig. 8.11.

However, the same sentence (fig. 8.10c) can also be thought of as an extension vis-à-vis the only instantiation of 8.10b (namely, 8.10a), as depicted in fig. 8.12.

Hence, both Soft_Constructional_Fails and successful derivations can be accounted for by a UBA: the former are extensions vis-à-vis strings previously encountered, the latter are instantiations of schemas that are classified as available units under Method_A.

---

[51] In this sense, the weak transitive schema is a schematic form-meaning pairing which is on its way to becoming entrenched enough to gain status of unit.

[52] For the meaning of the construction, refer back to examples (28) and (29) in 1.3.5.

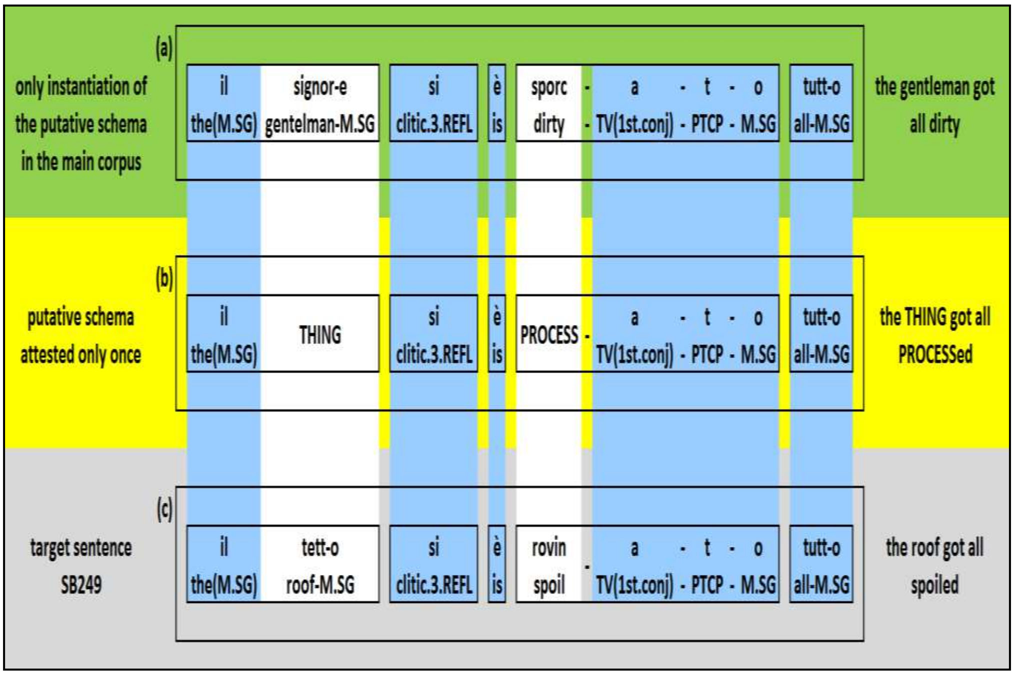


Figure 8.10: deriving target sentence SB249 (c) from its component units under Method_C and Method_D; 1 of 2. Slot formation is highlighted in white and recurring lexical material is highlighted in blue.

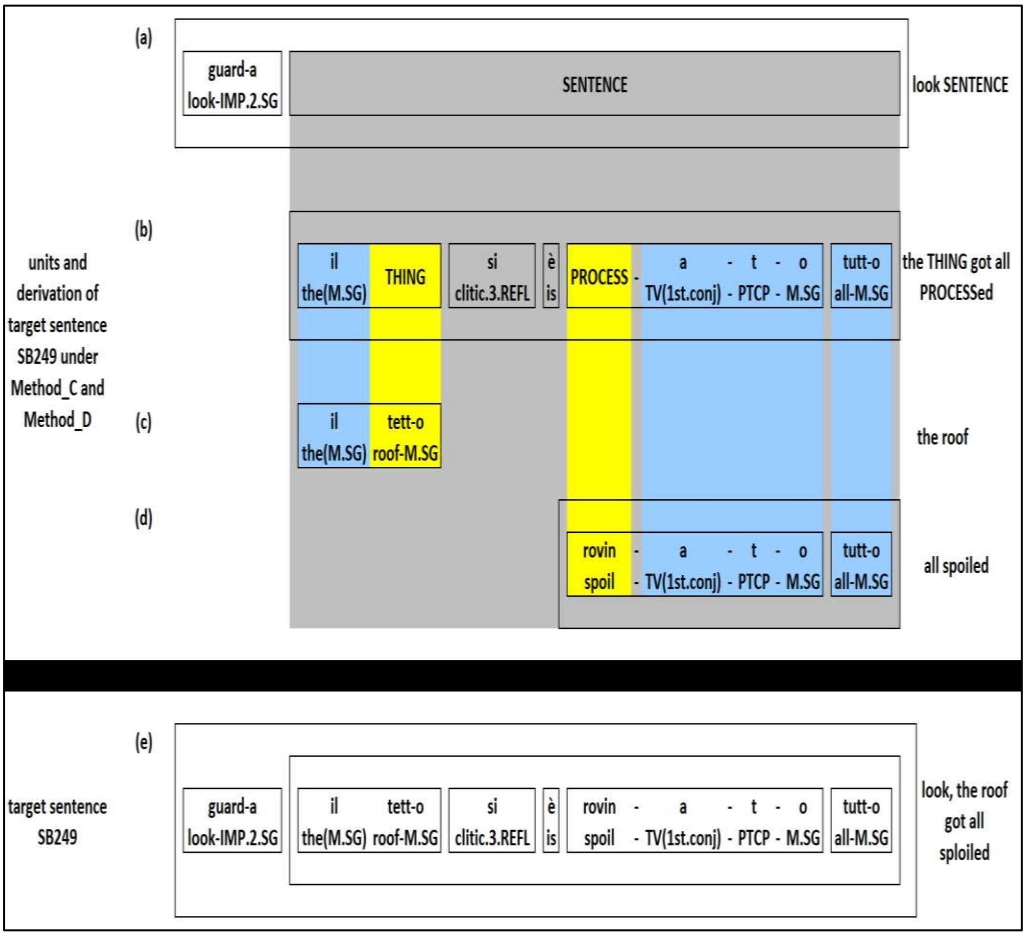


Figure 8.11: deriving target sentence SB249 (e) from its component units (a-d) under Method_C and Method_D; 2 of 2. Superimpositions are highlighted in grey and yellow. Shared lexical material is highlighted in blue.

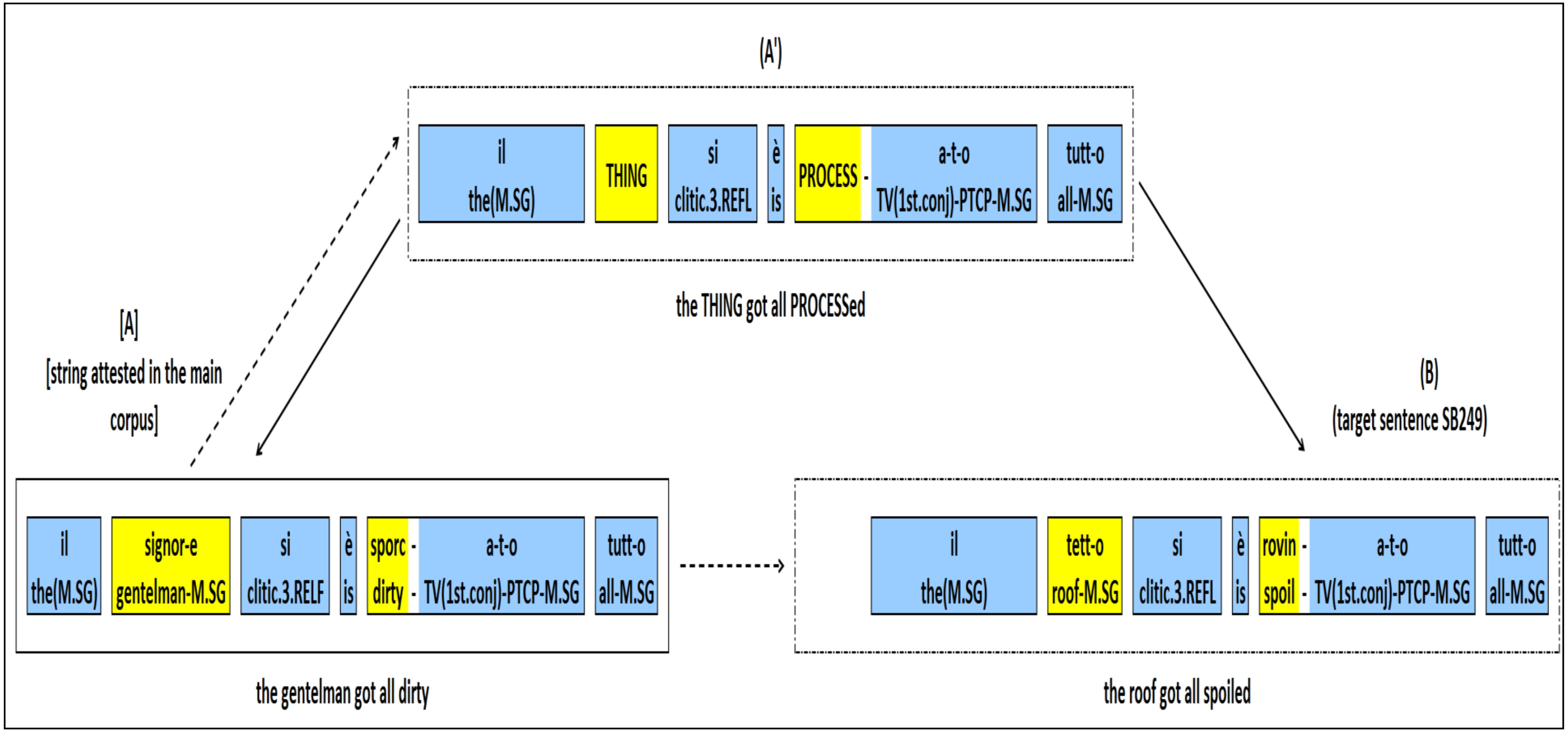


Figure 8.12: target sentence SB249 (B) as an extension vis-à-vis a string attested in the main corpus (A). Strings in thick boxes have status of units; strings in dashed boxes do not have status of unit. Solid arrows are relationships of elaboration; dashed arrows are relationships of extension. Words highlighted in blue indicate shared lexical material, the material in yellow indicate relationships of elaboration or extension.

8.3.3. ELABORATIVE DISTANCE AND ERROR RATES

Both extension and elaboration are cognitive processes, which therefore carry a processing burden with them. The two differ in what Langacker (2008) calls **elaborative distance**. Let us indulge in a simplistic metaphor and picture the "difference" between a model (SS) and a new sentence (TS) as a path to be walked through (from SS to TS). The path from a schema to its instantiation is often minimal, if only because a schema is already immanent in its instantiations; it is a (relatively) short and (very) straight walk. Conversely, apprehending B (TS) as a case of A (SS) through extension involves looking at B from A and then working out a way to it (creating the fleeting superordinate structure A'). Hence, extension is a path that is not straight and also longer than the one involved in elaboration.

I would then like to posit that, all things being equal, extension is a more demanding cognitive activity than elaboration, as it involves the "ad-hoc", fleeting abstraction of a superordinate structure. In the case of children, such a generalisation is also made on the basis of a linguistic competence that is both more concrete (lexically-specific) and based on a much poorer linguistic experience than that of adults. If this is true, those sentences that can be derived by positing extension processes should be more likely to yield ungrammatical, non-adultlike outcomes than derivations from fully instantiated schemas. The former derivations require a more demanding cognitive process and are based on less straightforward models (SS), which are less compatible with their TSs.

Indeed, Soft_Constructional_Fails are 9.8 times (odds ratio 3.05 – 29.55) more likely to present ungrammatical sentences than successful derivations (Fisher's Exact Test, p(two-tailed)< 0.0001; fig. 8.13).

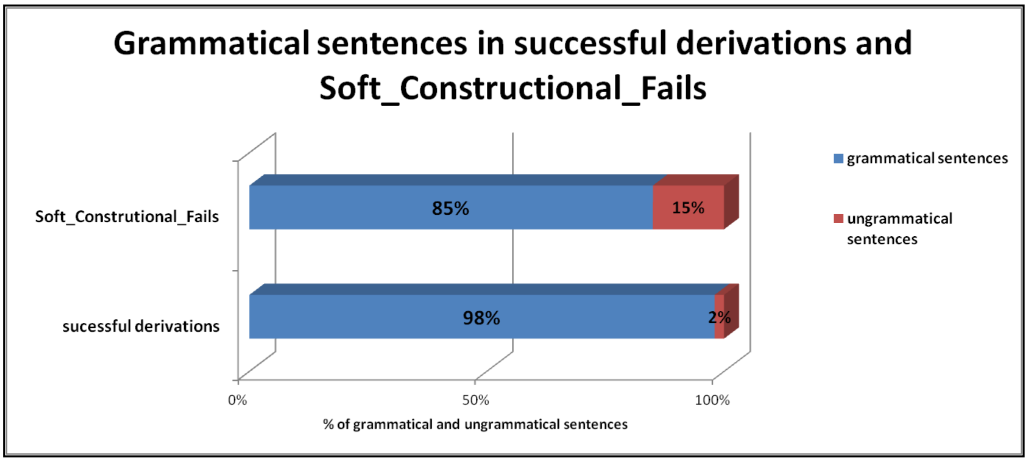


Figure 8.13 grammatical and ungrammatical sentences in successful derivations and Soft_Constructional_Fails.

However, the majority (85%) of Soft_Constructional_Fails are grammatical, well-formed sentences. This implies that, overall, when some sort of model is attested, Roberto, more often than not, draws the appropriate generalisations. Although this is undoubtedly true, a closer look at those utterances shows that target sentences that were classified as Soft_Constructional_Fails differ from their (putative) models only minimally. For instance, target sentence SA141 (54) is a four-operation derivation under both Method_C and Method_D. However, it differ from the closest string attested in the main corpus (55) in only two respects: the possessive *mia* (*mi-a*; "my-F.SG") is added before *nonna* "grandma" and the *THING_GIVEN* slot is instantiated by a different NP.

54. *La mia nonna mi ha dato questo piatto*
l-a *mi-a* *nonn-a* *mi* *ha*
the-F.SG my-F.SG grandma-F.SG clitic.1SG.DAT has
d-a-to *quest-o* *piatt-o*
give-TV(conj.I)-PTCP this-M.SG dish-M.SG
"My grandma gave me this dish."
(week.6.2014.02.18.14; line 1155)

55. *La nonna mi ha dato un boccon-e grand-e*
l-a *nonn-a* *mi* *ha*
the-F.SG grandma-F.SG clitic.1.SG.DAT has
d-a-to *un* *boccon-e* *grand-e*
give-TV(conj.I)-PTCP a(M.SG) mouthful(M.)-SG big-SG
"Grandma gave me a big mouthful."
(week.6.2014.02.18.14; line 1155)

Indeed, ungrammaticality seems mostly to arise in those sentences that can be qualitatively interpreted as being more distant from their models (as shall be discussed in section 8.4.3).

8.3.4. ACTIVATION SET AND SCHEMA'S SPECIFICATIONS

In section 2.4.3 it was argued that when speakers need to express themselves, they activate a set of units that compete with each other to categorise the target sentence. A trade-off between degree of overlap (specificity) and entrenchment determines which units are selected to categorise the new target expression. When two units that have conflicting specifications are selected, the specifications of either override the specifications of the other. A typical example of this is the formation of syntactic questions in English (refer back to section 2.4.3.3).

Two sentences that are classified as Soft_Constructional_Fails cannot be derived because their putative precedents present contrasting WO specifications. Those two sentences are traced back to *fixed-schemas* whose WO (e.g. *SBJ-COP-ADV-ADJ*) does not match exactly the linear WO of the target sentences (e.g. *COP-ADV-ADJ-SBJ*). However, those very same sentences are also traced back to other (needed) schemas whose WO matches the order attested in them. In the following paragraphs, it is argued that both those sentences can be explained by the fact that the WO specifications of one schema are overridden by the WO specifications of another schema whose WO matches the one attested in the target sentence.

Target sentence SB192 (56) presents the sequence *COP-ADV-ADJ-SBJ* and can be traced back to two schemas:

a) the schema in fig. 8.14a (yellow) - which presents a fixed and different WO (*SBJ-COP-ADV-ADJ*)

b) the schema in fig. 8.14b (yellow) - whose *ADV-ADJ-NOUN*[53] sequence matches the WO of the target sentence.

[53] Attested with and without copula.

56. *Era molto scarica la moto*
era molto scaric-a l-a mot-o
was very out.of.battery-F.SG the-F.SG motorbike(F.)-SG
"The motorbike was really out of battery."
(week6.2014.02.18.B: line 1491)

The interaction of the two schemas accounts for gender and number agreement. The schema in fig. 8.14a (yellow) accounts for subject-verb agreement and the schema in fig. 8.14b (yellow) accounts for gender-number agreement between adjective and NP. The WO specifications of the latter override the WO specifications of the former.

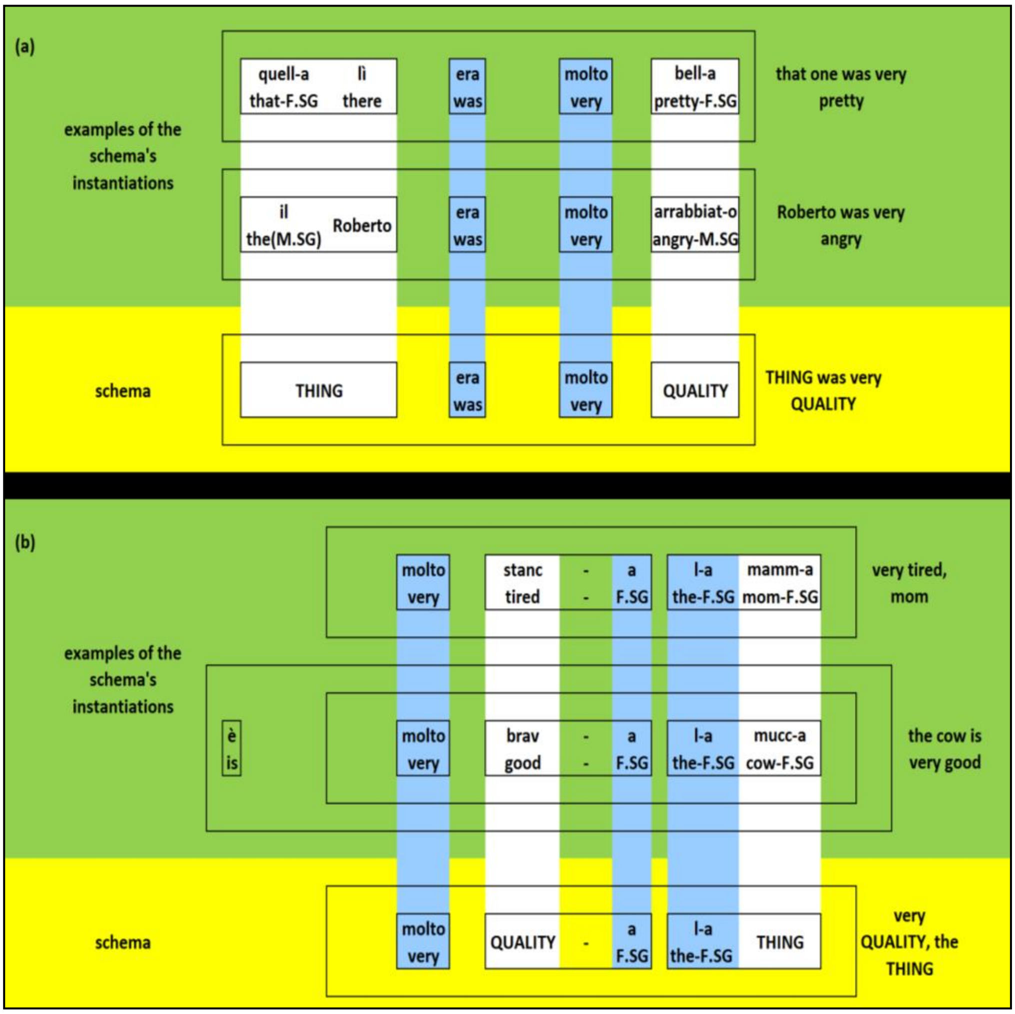


Figure 8.14 the precedents of target sentence SB192 (56). Schemas are in the yellow strips and their instantiations in the green strips. Relationships of elaboration are in white and shared concrete material is in blue.

Overall, although Method_A cannot strictly account for Soft_Constructional_Fails, a cognitive, UB approach to LA can provide a coherent interpretation of them, as well as of Roberto's morphological productivity and of why Soft_Constructional_Fails present a higher proportion of ungrammatical sentences than successful derivations.

8.4. HARD CONSTRUCTIONAL FAILS

Hard_Constructional_Fails are those target sentences that are classified as fails under all methods. There are fifty sentences that contain at least one such fail, and they represent about 7% of Roberto's dataset (768 target sentences). Those fails are all at the clause level (i.e. they are **syntactic fails**) and can be grouped into subcategories based on the issues that prevented their (successful) derivation (table 8.2).

Table 8.2: distribution of Hard_Constructional_Fails.

Type of Hard_Constructional_Fail	no. of sentences	% of 50 hcf	% of 768 target sentences
successful derivations that do not account for some kind of agreement within the sentence	13	26%	2%
accountable for by concatenations of partial overlaps	12	24%	1.6%
multiple issues	10	20%	1.3%
phonological mistake, element dropped from an attested schema and ill-formed imitation of the preceeding sentence	9	18%	1%
accountable for by insertion of a component unit within another component unit	5	10%	1%
explainable through extension and the co-activation of two different schemas	1	2%	0.1%
total	**50**	**100%**	**7%**

8.4.1. PHONOLOGICAL MISTAKES, OMISSIONS AND ILL-FORMED IMITATIONS

Nine sentences are classified as Hard_Constructional_Fails because Roberto mispronounces a word (four sentences), drops an element (e.g. a preposition) from a well attested schema (four sentences) or badly imitates an adult's utterance (one sentence). Phonological mistakes are attested when Roberto seems to mispronounce a word, these mispronunciations happen to be existing words (e.g. /ma/ "but" is realised as /da/ "from"). As a result, such target sentences cannot be traced back[54].

[54] They could have been traced back successfully if Roberto had pronounced the putative targets correctly.

Target sentence SB431 (57b) is probably the most interesting case of this sub-type of Hard_Constructional_Fails, as it stems from a failed attempt to elaborate on what the researcher had just said. Such a failed attempt results in Roberto using a locative instead of an adjective (underlined).

57. a) *RES: *io c'* ***ho la testa sottosopra****.*
io c(i) ***ho*** ***l-a*** ***test-a***
I clitic.DAT **have(PRS.1.SG)** **the-F.SG** **head-F.SG**
sotto+sopra
under+over
"I've got my head upside down."

b) *CHI: *hai la testa sopra.*
hai *l-a* *test-a* *sopra*
have(PRS.2.SG) the-F.SG head-F.SG over/up
"You've got the head over/up."

(week6.2014.02.18.B: lines 3098 and 3099)

In (57b) Roberto reuses the part in bold of the researcher's sentence (57a) and correctly conjugates the verb *avere* "to have": *ho* "have(PRS.1.SG)" becomes *hai* "have(PRS.2.SG)". However, he reuses only the second part of the compound *sottosopra* (/sot:o'zopra/, "upside down"; *sotto* "under" + *sopra* "up/over"). In the resulting sentence, an adverb of place (*sopra*; "over/up/on") is used instead of the adjective (*sottosopra*; "upside down") and Roberto's sentence cannot be traced back.

All but one of these nine sentences are ungrammatical. Because of their phonological and omission-rooted nature, these fails are not further analysed. However, the reader may find examples of omissions and phonological mistakes (and their description/analysis) in Appendix_II (ch. 20.3.2).

8.4.2. PROPORTION OF UNGRAMMATICAL SENTENCES AMONGST HARD_CONSTRUCTIONAL_FAILS

If the previously discussed nine sentences are taken out of the equation, there are forty-one Hard_Constructional_Fails; of these, fourteen (34%) are ungrammatical and twenty-seven (66%) are grammatical. Hence, such fails display a proportion of ungrammatical sentences that is more than twice as high as Soft_Constructional_Fails and seventeen times as high as successful derivations (fig. 8.15). Table 8.3 also shows that ungrammatical sentences are fairly equally distributed across Hard_Constructional_Fails, Soft_Constructional_Fails and successful derivations.

Table 8.3: distribution of ungrammatical sentences.

type of target sentence	no. of ungrammatical sentences	% of 41 ungrammatical sentences
successful derivations	11	27%
lexical fails	1	2%
Soft_Constructional_Fails (scf)	7	17%
Hard_Constructional_Fails (hcf)	14	34%
Phonological mistakes, elements dropped from an attested schema and ill-formed imitations.	8	20%
total	41	100%

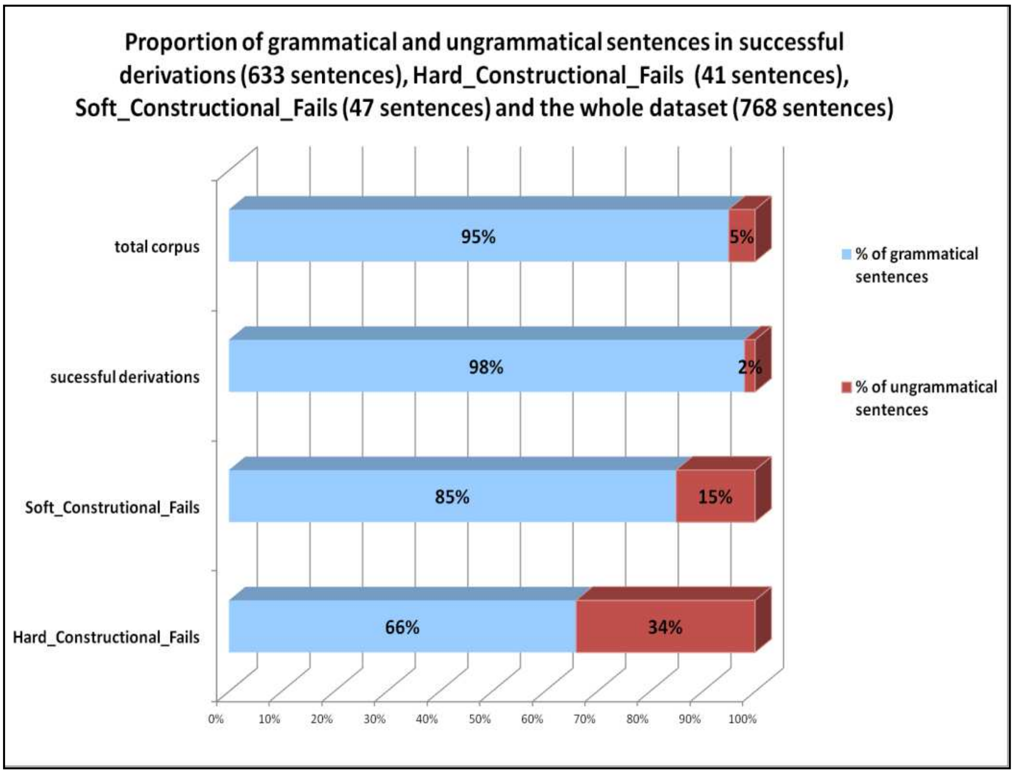


Figure 8.15: proportion of grammatical and ungrammatical sentences in successful derivations, Hard_Constructional_Fails, Soft_Constructional_Fails and the whole dataset.

If compared with (two-tailed) Fisher Exact Tests (corrected for Bonferroni adjustment, table 8.4) both Hard_Constructional_Fails and Soft_Constructional_Fails differ statistically from successful derivations in terms of the proportion of grammatical vs. ungrammatical sentences. The two types of fails do not differ statistically after Bonferroni adjustment (they do before, though: p=.046).

Table 8.4: comparing grammatical vs. ungrammatical sentences in Successful Derivations, Hard_Constructional_Fails and Soft_Constructional_Fails. The table presents odds ratios (and CIs) for each comparison.

	Successful Derivations	Soft_Constructional_Fails	Hard_Constructional_Fails
Successful Derivations			
Soft_Constructional_Fails	9.82** (3.05 - 29.55)		
Hard_Constructional_Fails	28.82** 11.04 - 77.77	2.93 0.95 - 9.77	

** p < .05 (Bonferroni correction)

Fig. 8.15 shows that the child overwhelmingly produces grammatical sentences (95%), which is in line with a well attested finding that children commit very few errors when they speak. The fact that successful derivations are overwhelmingly grammatical (98%) confirms that reliance on lexically-specific schemas and on input-based narrow generalisations helps children to produce adultlike sentences (and the occasional overgeneralisations, like (51) discussed in 8.1). However, when children have experienced few (Soft_Constructional_Fails) or very few (Hard_Constructional_Fails) lexically-specific patterns from which they could "build their sentences together", they are far more likely ($p < .05$) to produce non-adultlike sentences (see Lieven & Brandt (2011) for a similar argument).

In the next few subsections I will attempt to give an insight into Hard_Constructional_Fails and what they suggest about Roberto's linguistic representation.

8.4.3. FUNCTIONAL COERCION, EXTENSION AND CO-ACTIVATION OF COMPETING UNITS

In section 8.3.3, it was stated that when a model (SS) is used to categorise by way of extension a target structure (TS) which appears to be qualitatively distant from it, ungrammatical sentences are more likely to arise. In section 8.3.4 it has been further argued that when units with conflicting specifications are activated, the specifications of either one override the specifications of the other.

Target sentence SB473 (58) and target sentence SB468 (59) can be accounted for by the interaction of:

a) the co-activation of two units in which neither unit imposes its specifications on the other,

b) non-adultlike generalisations based on an extension from a prototype.

58. *CHI: *tu <sei> [/] <sei> [/] sei come (..) da tagliare la carne, Luca.*
tu sei come da tagliare la carne, Luca
tu *sei* *come* *(..)* *da*
you(2.SG.NOM) be(PRS.2.SG) like (..) to
tagli-a-re *l-a* *carn-e* *Luca*
cut-TV(conj.I)-INF the-F.SG meat(F.)-SG Luca
"Luca, you are like (..) to cut the meat."
(week6.2014.18.B: lines 3372-3373)

59. *CHI: **devi dare il gistratore [: registratore] [* p] a qualcuna [: qualche] [*] persona.*
devi dare il registratore a qualcuna persona.
dev-i *d-a-re* *il*
have.to-PRS.2.SG give-TV(conj.I)-INF the(M.SG)
registrator-e *a* *qualcun-a* *person-a.*
voice.recorder(M.)-SG to someone-F.SG person-F.SG
("*you have to give the voice recorder to someone person").
"You have to give the voice recorder to someone."
(week.6.2014.02.18.B: line 3324-3325)

SB468 (59) belongs to those sentences that present multiple issues (table 8.2). The part under analysis now is the string **qualcuna persona* "someone person". Roberto uses a pronoun (*qualcun-a*, "someone-F.SG") instead of an adjective (*qualche*, *"some")*. Essentially, he seems to co-activate the two schemas in (60).

60. a) qualche THING-a.
some THING-F.SG.

b) *qualcun-GENDER.NUMBER.*
someone-GENDER.NUMBER.

This ungrammatical use of the pronoun instead of the adjective is attested once in the child's own main corpus (61). Note, furthermore, that *qualcun-**a*** "someone-F.SG" in (59) is also a morphological fail, as only the masculine form (*qualcun-**o***) is attested in the main corpus.

61. *CHI: _*<respondiamo [: rispondiamo] [* p] > [<] a qualcuno [: qualche] [* s:r]> [?] bambido [: bambino] [* p]_
rispondiamo a qualcuno bambino

rispond-iamo	_a_	_qualcun-o_	_bambin-o_
answer-PRS.1.PL	to	someone-M.SG	child-M.SG

("*we answer to someone child").
"We answer to some child."
(week5.2014.02.04: lines 1389-1390)

Roberto does not seem to fully distinguish between the pronoun (*qualcuna*, /kwual'ku.na/) and the adjective (*qualche*, /'kwual.ke/). The attempted (grammatical) target is likely to be *qualche persona* (literally *"some person"*), for which the method identified the pattern *qualche THING-a* "some THING-f.sg" (60a).

Constructions are form-meaning pairings and children develop schematic patterns by analogising on the basis of formal and functional similarities across concrete strings (*I kick you, we kick the ball*) and, later on in development, across schemas (*KICKER kick KICKEE*; *HITTER hit HITTEE*). Such use of the pronoun instead of the adjective seems to be rooted in a generalisation based on phonological and semantic similarities between *qualche* (/'kwual.ke/ "some") and *qualcuna* (/kwual'ku.na/ "someone-F.SG"). The two words share the sequence of phonemes /kwualk/ (formal similarities) and both indicate an indefinite quantity of a possibly indefinite entity (semantic similarities). It seems that Roberto tries to use the pronoun in a construction in which he heard only the adjective (*qualche THING-a*; "some THING-f.sg"). What might be going on is that Roberto assimilates the pronoun to the adjective on the basis of shared semantic and phonological features. As a following-up step, he extends the distributional patterns of the latter to the former. To some extent, it could be argued that *qualche THING-a* "some THING-f.sg" (60a) represents the prototype – in Goldberg's (2006) terms - from which a construction *X THING* is inferred. X indicates some unspecified quantity of a particular entity (*THING*). Since the "indefinitiveness" of *qualcuno* (pron.) somehow matches the "indefinitiveness" of *qualche* (adj.), the distributional patterns of the latter are extended to the former (i.e. *qualcuna-GENDER.NUMBER* is used to fill the slot *X* of *X THING*).

Target sentence SB473 (58) can be interpreted in several ways. The interpretation of *come* “like” is crucial. *Come* can be used as an interjection in the same fashion as *like* in English. This interpretation could be supported by the pause between *come* and the rest of the sentence. The same pause however, could be interpreted as evidence that Roberto is co-activating three constructions (fig. 8.16a-c).

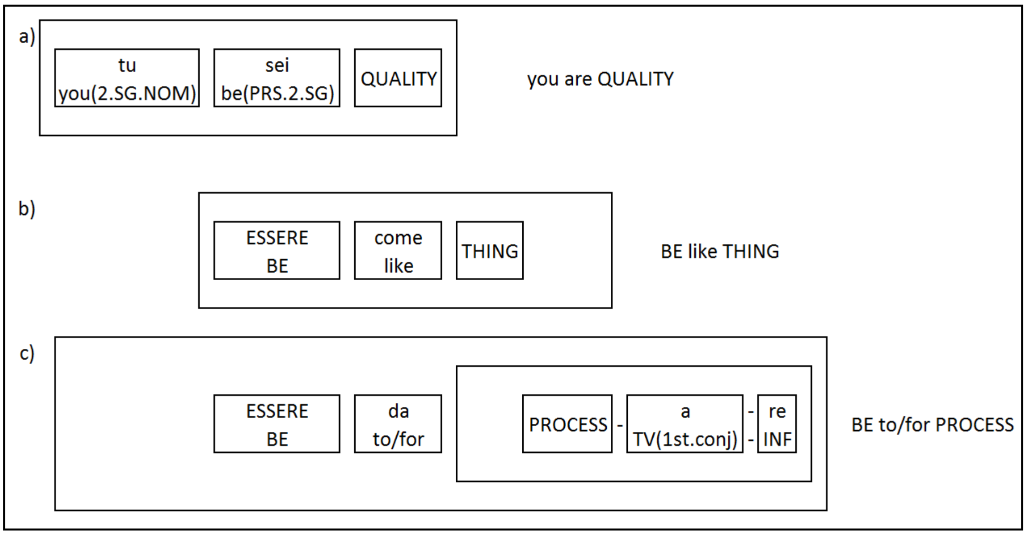


Figure 8.16: (some of) the putative precedents of target sentence SB473 (58).

The hypothesis I would like to put forward is that Roberto interprets *da tagliare la carne* “for cutting the meat/to cut the meat” as a *THING* and uses it to fill the slot *THING* of fig. 8.16b, as illustrated in fig. 8.17.

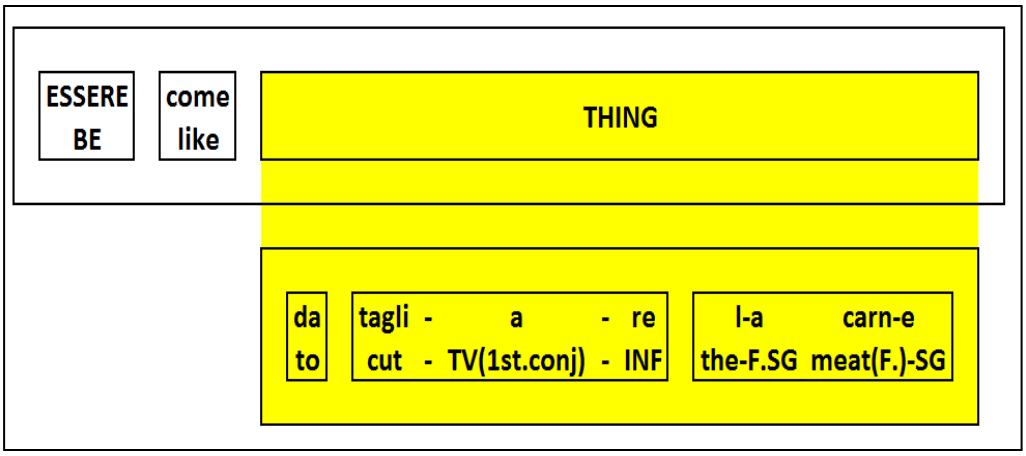


Figure 8.17: elaborating a *THING* slot with *da tagliare la carne*. The superimposition is in yellow.

Why would Roberto interpret a string that clearly instantiates a *PROCESS* as a string instantiating a *THING* (why would he think of a VP as an NP)?

Because he categorises the TS *da tagliare la carne* "for cutting the meat/to cut the meat" as an extension vis-à-vis *da mangi-are* "to eat-INF", which can profile a *THING* (62).

62. [*da mangiare*] -----► (*da tagliare la carne*).

Da mangiare "to eat-INF" is an instantiation of the schema in (63) which represents a *PROCESS* governed by a PP, very much like in English (64a). However, this construction can also be used to profile a *THING* – in Langacker's (2008) terms. In this NP-like usage, the meaning of the construction can be paraphrased as "*something to PROCESS"* (*da mangiare*, "to eat"; "something to eat") or *"somewhere to PROCESS"* (*da dormire* "to sleep"; "somewhere to sleep")*,* depending on the semantics of the verb instantiating the *PROCESS slot*. Such usage or "profiling" is mainly found in informal speech and is mainly attested when the construction is used in combination with certain verbs of transfer (such as *give* and *bring*) and the verb *fare* "to do". Instances of such NP-like usage of the construction typically involve verbs that are nutrition-related (64b-c, in **bold**)

63. da PROCESS-TV-re.
to PROCESS-TV-INF.
"To PROCESS/ for PROCESSing."

64. a) *RES: *il tacchino è buono da mangiare*
il *tacchin-o* *è* *buon-o*
the(M.SG) turkey-M.SG is good-M.SG
da *mangi-a-re*
to eat-TV(conj.I)-INF
"The turkey is good to eat."
(week2.2014.01.14.A: line 384)

b) *MOT: *serve per dare* ***da bere*** *ai fiorellini .*
serv-e *per* *d-a-re*
be.for-PRS.3.SG to give-TV(conj.I)-INF
da ***b-e-re*** *a=i*
to **drink-TV(conj.II)-INF** to=the(M.PL)
fiorell-in-i
flower-little(DIM)-M.PL
("(It) is for giving the little flowers **(something) to drink**").
"It is for watering the flowers."
(week2.2014.01.20: line 423)

c) *CHI: *ti do* ***da mangiare***
ti *d-o* ***da*** ***mangi-a-re***
clitic.2.SG.DAT give-PRS.1.SG **to** **eat-TV(conj.I)-INF**
"(I) give you (**something) to eat**."
(week4.2014.01.31: line 932)

Da mangiare "to eat" can be thought of as the prototype of this NP-like usage of the construction and can be paraphrased as *food* or *something to eat*. Across the whole dataset the string *da mangiare* is attested sixty-three times, of which fourty-eight (71%) are instantiations of an NP-like usage (like in 64c). This suggests that *da mangiare* is more frequent (hence, more entrenched) as a string profiling a *THING* than as a string profiling a *PROCESS* (χ^2 (1)= 4.81, odds ratios = 2.4 (1.09 – 5.42), *p*(two-tailed) < .02, if a 50-50% chance level is assumed).

A further factor to consider is that the infinitive suffix can be used to derive de-verbal nouns (65).

65. Sputare non è bello
sput-a-re *non* *è* *bell-o*
spit-TV(conj.I)-INF not is nice-M.SG
"Spitting is not nice."

The interaction of a few factors might lead Roberto to apprehend *da tagliare la carne* as an extension vis-à-vis *da mangiare*:

a) De-verbal nouns can be derived by using the infinitive form and both *da* ***mangiare*** and *da* ***tagliare*** can be thought of as de-verbal nouns (like *sputare* "to spit" or "spitting" in 65).

b) *Da mangiare* "to eat" and *da tagliare la carne* "to cut the meat" share morphological material; the preposition *da* and the infinitive marker –*are* (i.e. they both are instantiations of 63).

c) The schema in (63) can profile a *THING*, particularly when it comes to informal language.

d) *da mangiare* is often used as a signifier for food (hence it can profile a THING).

e) Both strings are food-related (one cuts the meat before eating it).

As an extension vis-à-vis *da mangiare*, *da tagliare la carne* inherits its profile and it is used to fill the slot *THING* of fig. 8.16b (see fig. 8.17).

According to the interpretations given to (58) and (59), two observations can be made.

Firstly, these sentences could be seen as evidence that Roberto's syntactic categories are not adultlike and that he is developing them by drawing generalisations from the concrete language he hears and the lexically-based patterns it provides. Indeed, these two examples cast doubt on the ability of Roberto to assign words to syntactic categories and hence to parse the input in an adultlike manner. They are not isolated cases; another similar case of misinterpretation of a particular word's syntactic category is his usage of *nevic-a* "snow-PRS.3.SG" instead of the noun *nev-e* "snow(F.)-SG".

Secondly, (58) and (59) support the hypothesis that when children try to go beyond what they know (beyond their inventories of lexically specific patterns) they are likely to utter ungrammatical (non-adultlike) sentences.

8.4.4. AGREEMENT ISSUES

Across the whole dataset, there were twenty-six target sentences[55] (3%) that **could** be derived by applying superimposition and/or juxtaposition to attested units *but* the interaction of their putative component units could **not** fully account for all the agreements involved. These sentences can be grouped into three main categories according to the grammatical phenomena involved:

a) Agreement between an object-NP and some other co-indexed resumptive elements (six clitic pronouns and a demonstrative pronoun); seven sentences.

b) Agreement between the subject and other elements in the sentence (such as adjectives and past participles); fifteen sentences.

c) Subject-verb agreement; four sentences.

Sentences of type (b) and (c) were classified as fails while those of type (a) were classified as successful derivations. The next sub-sections analyse these sentence types and the reasons for classifying them as either possible or non-possible derivations.

8.4.4.1. Object Agreement (seven successful derivations)

Method_A could not fully account for the agreement between an accusative clitic pronoun and either a right (fig. 8.18a) or a left dislocated object-NP in six sentences. Furthermore, in the target sentence depicted in fig. 8.18b, it was not possible to account for gender-number agreement between the DO *il registratore* "the(M.SG) voice recorder" and *quest-o* "this-M.SG". Five of the seven sentences were grammatical (as in fig. 8.18a-b), whereas two of them presented incorrect agreements (as in fig. 8.19a-b).

[55] Seven represents successful derivations (discussed in 8.4.4.1), six are classified as Soft_Constructional_Fails and thirteen as Hard_Constructional_Fails.

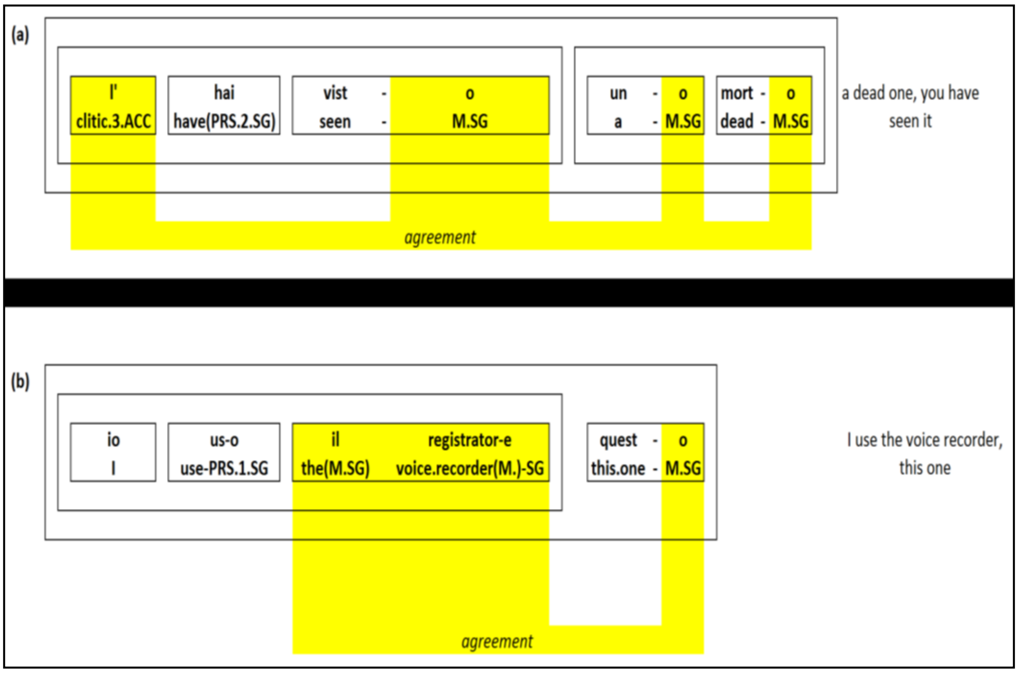


Figure 8.18: agreement between object-NML and resumptive element. Correct agreement is in yellow.

Figure 8.19: agreement between object-NML and resumptive element. Wrong agreement is in green.

These sentences could be accounted for by either:

a) a schema with a *THING* slot co-indexed with the resumptive clitic, which could be instantiated (elaborated) by an NP. Such a slot included information about number (*THING*). However, because it was instantiated in the main corpus by a wide range of NPs, it did not include any morphologically-specified fixed element that could provide information on gender (fig. 8.20a, which derives fig. 8.18a)

or by

b) juxtaposing the clitic-based sentence and the object-NP (fig. 8.20b, which derives fig. 8.19b).

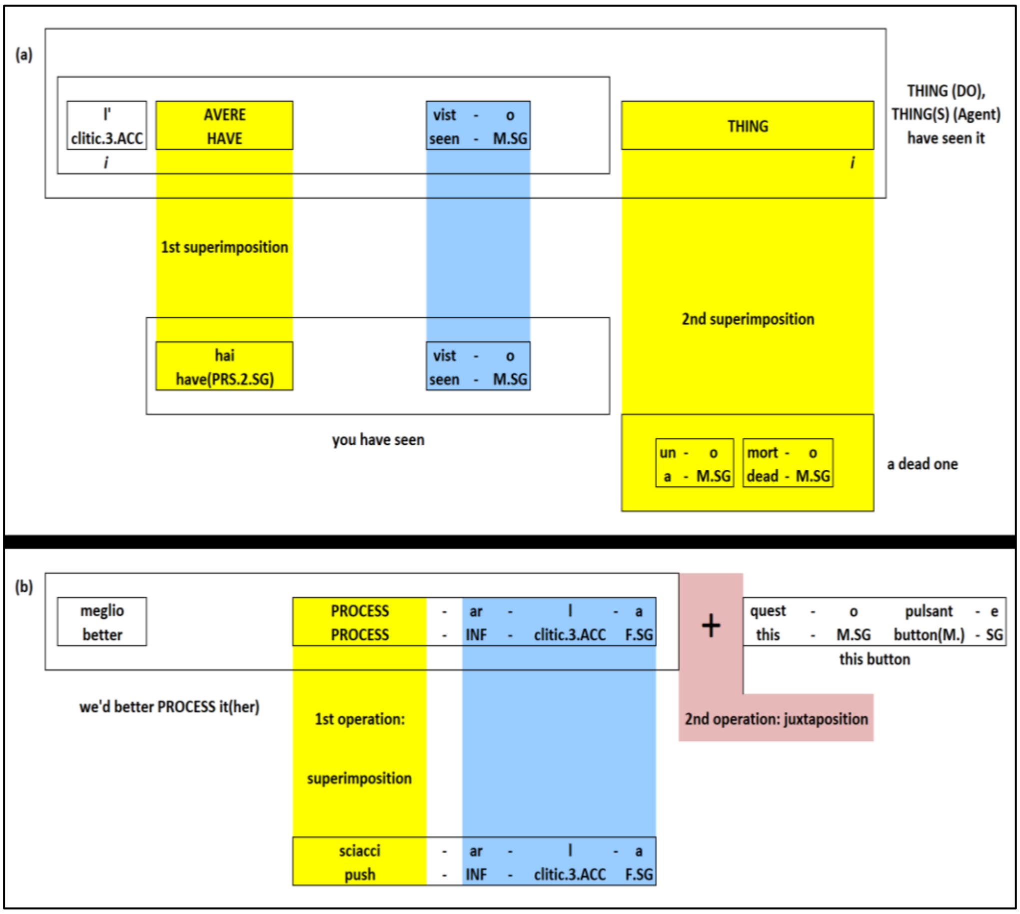


Figure 8.20: Deriving fig. 8.18a (a) and fig. 8.19b (b). Slot elaboration is highlighted in yellow and shared lexical material is in blue. The red highlighting and the sign plus (+) indicate the juxtaposition of units.

As the reader may recall (section 6.3.5), in order for juxtaposition to be allowed, no agreement had to be involved. Hence, a derivation like fig. 8.20b should not be allowed, as there should be agreement between the dislocated object-NP and the clitic pronoun (even if the child gets it wrong in this particular example). As for derivations like fig. 8.20a, it is also a derivation that should not be permissible. The schema does not specify the gender of the NP that can elaborate the slot. Hence, it is not clear on what basis the child does not use a feminine noun to fill the slot.

However, for those seven sentences, and for those seven sentences only, the method was relaxed and their derivations were considered successful. This decision was taken on the basis of three observations:

a) Those sentences can be thought of as topic-comment constructions, in which the object-NML and the clitic-based clause are two separate juxtaposed clauses, one having topic and the other having comment function.

b) If one assumes that point (a) is correct, then there are four possible outcomes: the resumptive element and the co-indexed NML agree in gender and number, they agree only in gender or only in number and, finally, they agree in neither gender nor number. This means that there is a 25% chance of producing the correct agreement[56]. Therefore, two agreements out of seven could be delivered correctly by chance alone. The child produced five correct and two wrong agreements. According to Fisher's Exact Test, this does not differ statistically from the chance level just hypothesised (odds ratio = 5.37(1.18- Inf.), *p*(one-tailed)=.14). Hence, it is at least questionable that Roberto has acquired competence of agreement between (dislocated) object-NMLs and other (co-indexed) elements.

c) More importantly, these constructions appear to be the linguistic environment in which adults participating in this study made most of their agreement mistakes[57]. Fig. 8.21a reports an instance of this construction uttered by the researcher, in which agreement is wrong (8.21b is the correct version of 8.21a).

[56] This is an underestimate, as in three cases a *THING* slot is attested. Hence, it is often about either masculine or feminine form, as the number is specified by the slot.

[57] This observation is made by casually inspecting the dataset. Whether this holds statistically is a question for another research.

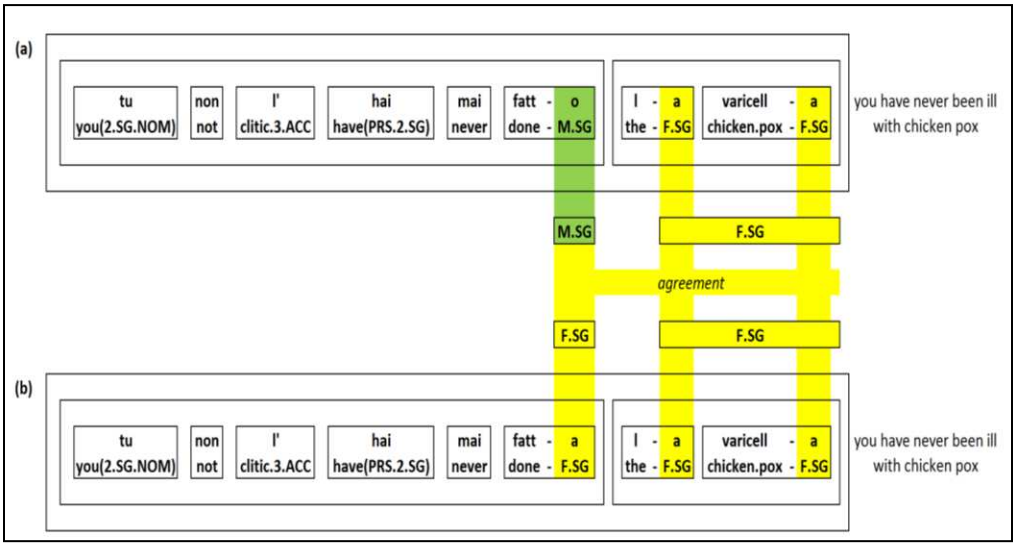


Figure 8.21: when adults deliver wrong agreement (highlighted).

Hence, because of the interaction of pragmatic (a), statistical (b) and input-dependent (c) factors, those seven sentences were considered successful derivations.

8.4.4.2. Agreement between subject and other phrasal constituents

Amongst the 768 target sentences, there were 15 utterances that presented agreement between the subject and other clasusal elements that could not fully be accounted for by using Method_A. This included agreement between the subject and a past participle (six sentences[58]; 66), an adjective (four sentences; 67), or an NP in what I would call the "explanative construction" (*THING_A is* *THING*_B or *THING_A is not THING_B*[59]; five sentences, 68)[60].

66. *CHI: *questo omino è rimasto lì dentro*
*quest-**o*** *om-in-**o*** *è* *rima-st-**o***
this-**M.SG** man-little(DIM)-**M.SG** is stay-PTCP-**M.SG**
lì *dentro*
there inside
"This little man stayed in there."
(week6.2014.12.18.A: line 988)

---

[58] One of these presented multiple issues that prevented its derivation.
[59] In this construction, the post-verbal NP is a *predicative complement of the subject*. According to Salvi (2001b), gender-number agreement between the two NPs is not always compulsory. In the following discussion, I consider such sentences as presenting correct agreement when either the agreement was correct or not grammatically required. However, these sentences were classified as fails because the instantiations of the schemas identified as their putative precedents could account for neither the specific form used by Roberto nor for the basis on which he might have inferred whether agreement was required or not.
[60] For the sake of brevity, I report only the relevant part of these examples. Agreement is emboldened.

67. *CHI: *gli stivaletti sono tutti sporchi*

gli	*stival-ett-**i***	*sono*	*tutt-**i***
the(M.PL)	boot-little(DIM)-**M.PL**	be(PRS.3.PL)	all-**M.PL**

*sporch-**i***
dirty-**M.PL**
"The little boots are all dirty."
(week6.2014.02.18.A: line 1100)

68. *CHI: *quello non è uno uovo*[61]

*quell-**o***	*non*	*è*	*un-**o***	*uov-**o***
that-**M.SG**	not	is	a-**M.SG**	egg-**M.SG**

"That is not an egg."
(week6.2014.02.14.B: line 1662)

These sentences, unlike the seven previously discussed, were considered fails because:

a) They cannot be interpreted as two clauses that are juxtaposed. Instead, agreement is undoubtedly within the same clause.

b) Adults do not seem to make many agreement mistakes in these types of constructions[62].

c) If all three constructions are considered together, there are nine correct agreements and six incorrect agreements. If one assumes a 25% chance level (three sentences out of fifteen), Roberto's rates of correct agreement are significantly (p(one-tailed)=.03) above chance (Fisher's Exact Test, odds ratio 5.61(1.18 – Inf)[63]).

However, five out of the nine correct agreements produced by Roberto are classified as Soft_Constructional_Fails. Thus, for those five sentences there are models attested in the main corpus that could account in lexically-specific terms for the basis on which Roberto produces correct agreements.

---

[61] Technically, in this sentence Roberto uses the wrong article, *uno uovo* instead of *un uovo. Un* should be used when the NP starts with a vowel. However, since both *un* and *uno* are indefinite masculine articles, the sentence is considered grammatical as the child appropriately chooses the M.SG article.

[62] Again, this is an observation that is made by casually inspecting the dataset.

[63] Note however, this is an underestimate as all schemas have a slot that specifies the number (*THING* or *THINGS*) of the filler. Hence, when needed, agreement is a matter of chosing between two forms (M. vs. F.), rather than four (M.SG, M.PL, F.SG and F.PL). This stricter threshold raises the chance level to 50-50%. In this case, Roberto's performance is not above chance (Fisher's Exact Test, odds ratio = 1.68 (0.4 – Inf.), p(one-tailed)=.36).

Two of those five sentences were target sentences that could not be derived because of conflicting WO specifications (as illustrated in sentence (56), fig. 8.14). The remaining three sentences could be interpreted as extensions vis-à-vis strings attested once or more in the main corpus. Such strings represent models on the basis of which the correct agreement could be delivered.

Interestingly, if those five grammatical sentences are excluded from this subgroup of fifteen sentences, six of the remaining ten sentences present incorrect gender-number agreement between the subject and another element. Hence, it would appear that when no (lexically-specific) model for agreement is attested in the main corpus strings, Roberto produces more wrong (six sentences) than correct (four sentences) agreements. If one assumes a 25% chance level for correct agreements (two out of ten sentences), Roberto's production of (correct) agreement is not above chance, not even before Bonferroni adjustment (Fisher's Exact Test, odds ratio = 5.43(0.79 – Inf), *p*(one-tailed)=.8).

This suggests that when there is no lexically-specific model providing a clear indication of correct agreement, Roberto is as likely to utter grammatical as ungrammatical sentences.

8.4.4.3. Subject-verb agreement

The analysis undertaken using Method_A cannot account for the agreement between the subject (il papà "the daddy") and the verb of the subordinate clause (faceva "was making") in target sentence SA071 (69). This sentence could be derived as in fig. 8.22.

69. *CHI: *quando c' era (.) il papà &*MOT:sì*[64] *che faceva (.) la pizza.*
quando c(i) er-a il
when there(clitic.LOC) be(IMPERF)-3.SG the(M.SG)
papà che fac-ev-a l-a
daddy(M.) that make-IMPERF-3.SG the-F.SG
pizz-a.
pizza-F.SG
"When there was daddy making pizza."
(week6.2014.02.18: line 683)

[64] Please note that, according to CHAT conventions (MacWhinney, 2000), the insertion of "&*MOT" means that the mother just says the word *sì* "yes", without interrupting the child's speech. That is, the mother encourages the child to continue, which he does by adding the relative clause to the main one.

The problem with the derivation of the above sentence lies in fig. 8.22c, in which the schema *faceva THING* (*was making THING*, vi) fills the slot *PROCESS* of (v). Although the profile of the former is compatible with the specifications of the slot of the latter, this operation is problematic because neither schema accounts for subject (*papà* "daddy") – verb (*faceva*; "was making") agreement. Since the *PROCESS* slot does not specify any morphological feature, it is not clear what prevents Roberto from filling the slot with a wrongly conjugated verb. The only derivations that are possible are the ones that account for *all* agreements attested. Since there is no other schema that can interact with *there was THING that PROCESS* to explain subject-verb agreement, this sentence is considered a fail.

The other three sentences that involve subject-verb agreement have a 2.SG pronoun as subject: two present correct subject-verb agreement and one presents an incorrect agreement.

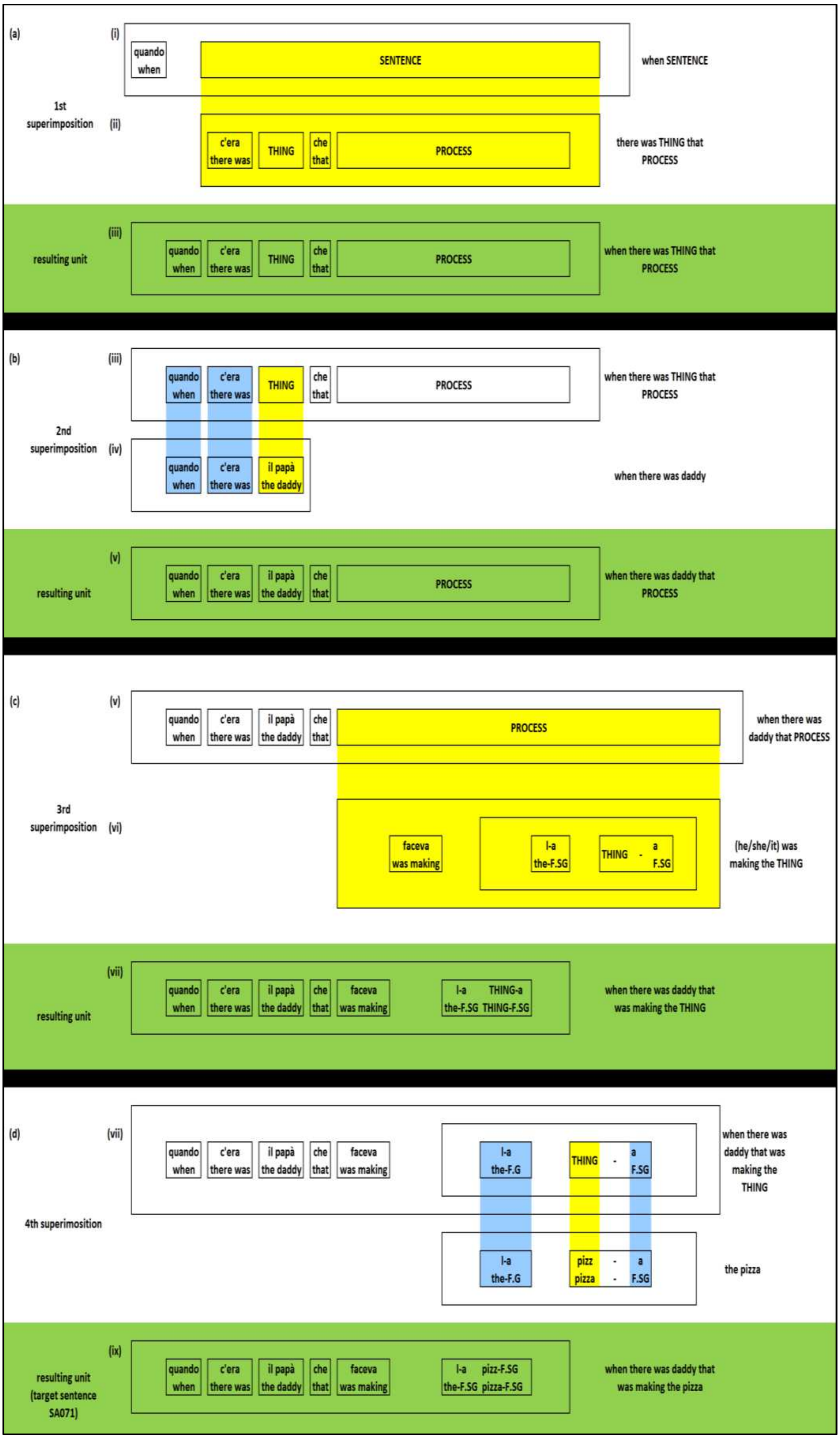


Figure 8.22: deriving target sentence SA071 (69). 8.22c is the problematic superimposition that causes the fail. Slot elaboration is highlighted in white and shared lexical material is highlighted in blue.

8.4.5. CONCATENATIONS OF PARTIAL OVERLAPS

Yet, there appears to be a way to account for target sentence SA071 (69) and explain what prevents the child from filling the slot *PROCESS* in *there was daddy that PROCESS* (fig. 8.22v) with a verb conjugated ungrammatically; by assuming a more flexible approach to superimposition. Namely, **concatenations of partial overlaps**, in which units are expanded by means of superimpositions (elaborations) of elements they share. Subject-verb agreement in SA071 can be accounted for by positing that fig. 8.22v is superimposed with fig. 8.23c, rather than with 8.22vi. The schema in fig. 8.23c specifies the verb inflection and therefore accounts for subject-verb agreement. This superimposition is illustrated in fig. 8.24. The string resulting from the superimposition of these two units (fig. 8.24c or 8.25a) is then superimposed (fig. 8.25) through partial overlap with fig. 8.22vi (unit b in fig. 8.25). The partial overlap is possible because 8.25b elaborates the slot of 8.25a and the two share the inflection *–ev-a*. In doing so, the partial overlap of the two units brings about the expansion of 8.25a that now includes a direct object (fig. 8.25c). How the whole sentence could be derived is illustrated in fig. 8.26.

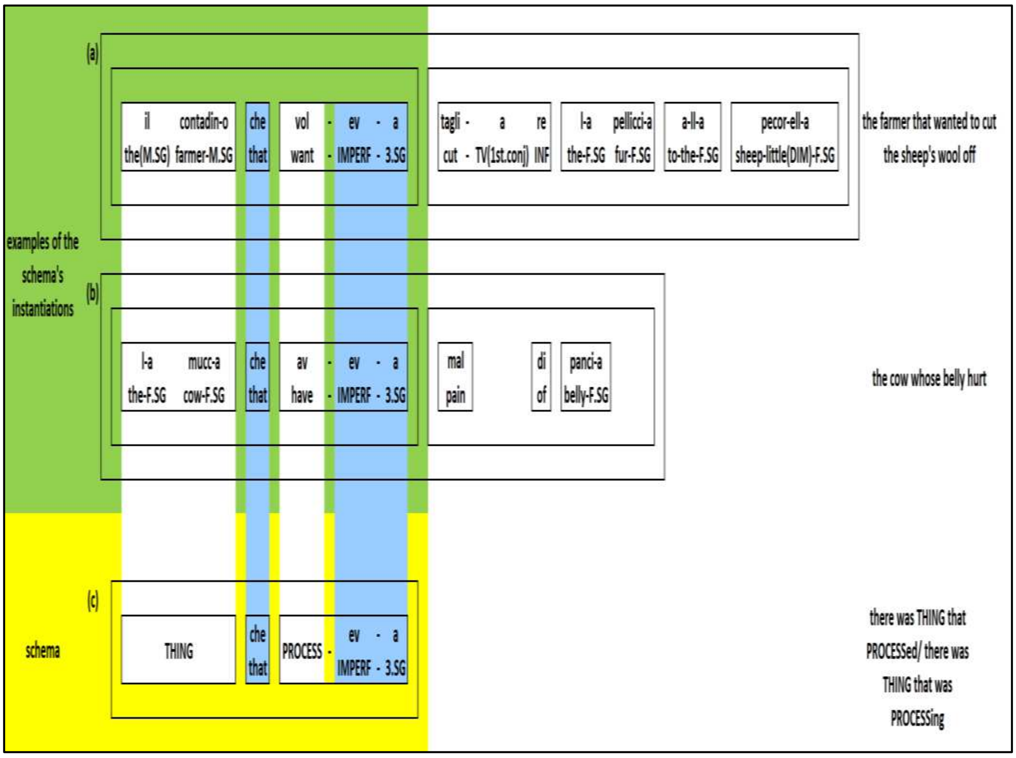


Figure 8.23: the schema (in yellow) that could account for subject-verb agreement in target sentence SA071 (69) and it instantiations (in green). The parts with white backgrounds do not contribute to the creation of the schema.

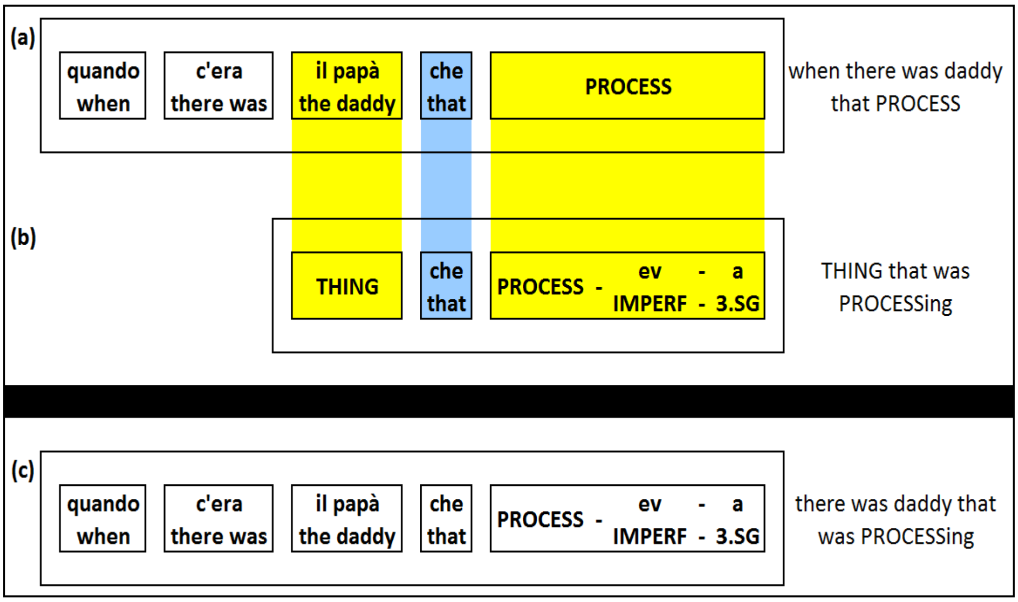


Figure 8.24: explaining subject-verb agreement in target sentence SA071 (69). The unit in c results from the superimposition of a and b. Slot elaboration is in yellow and overlapping (shared) lexical material is highlighted in blue.

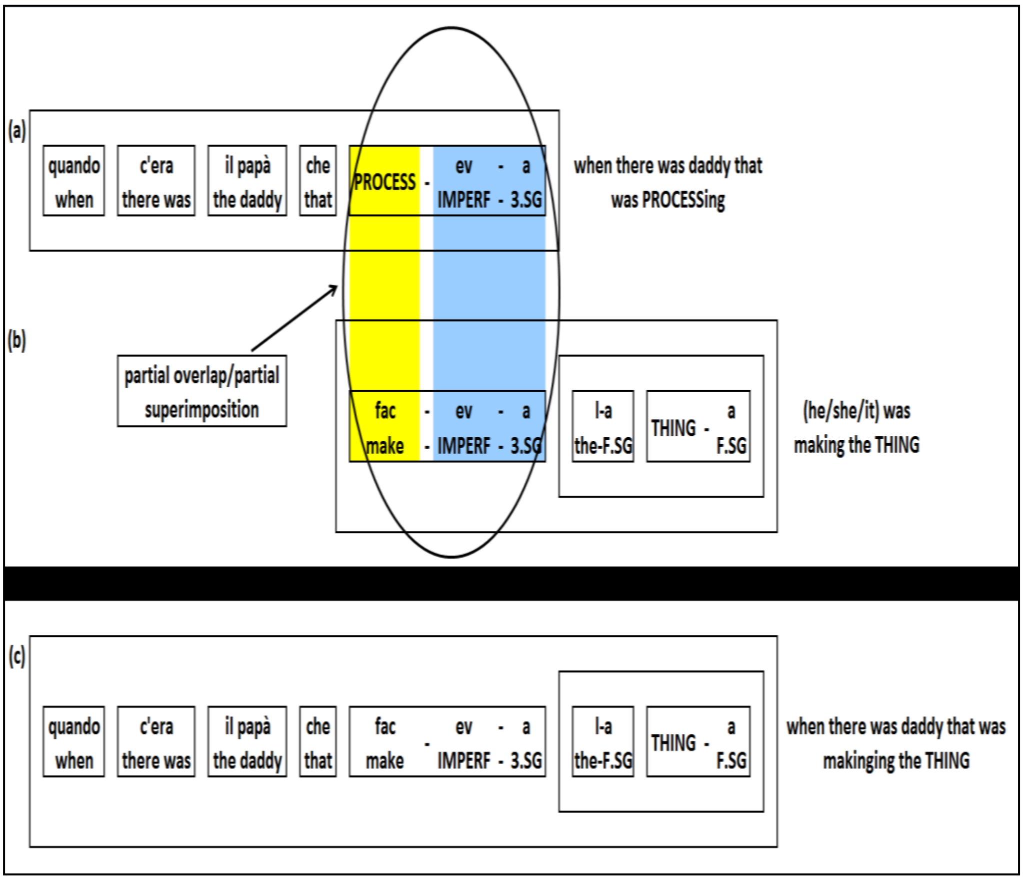


Figure 8.25: explaining target sentence SA071 (69) through partial overlap. The unit in c results from the partial overlap of a and b. See previous figure on how to interpret the colours.

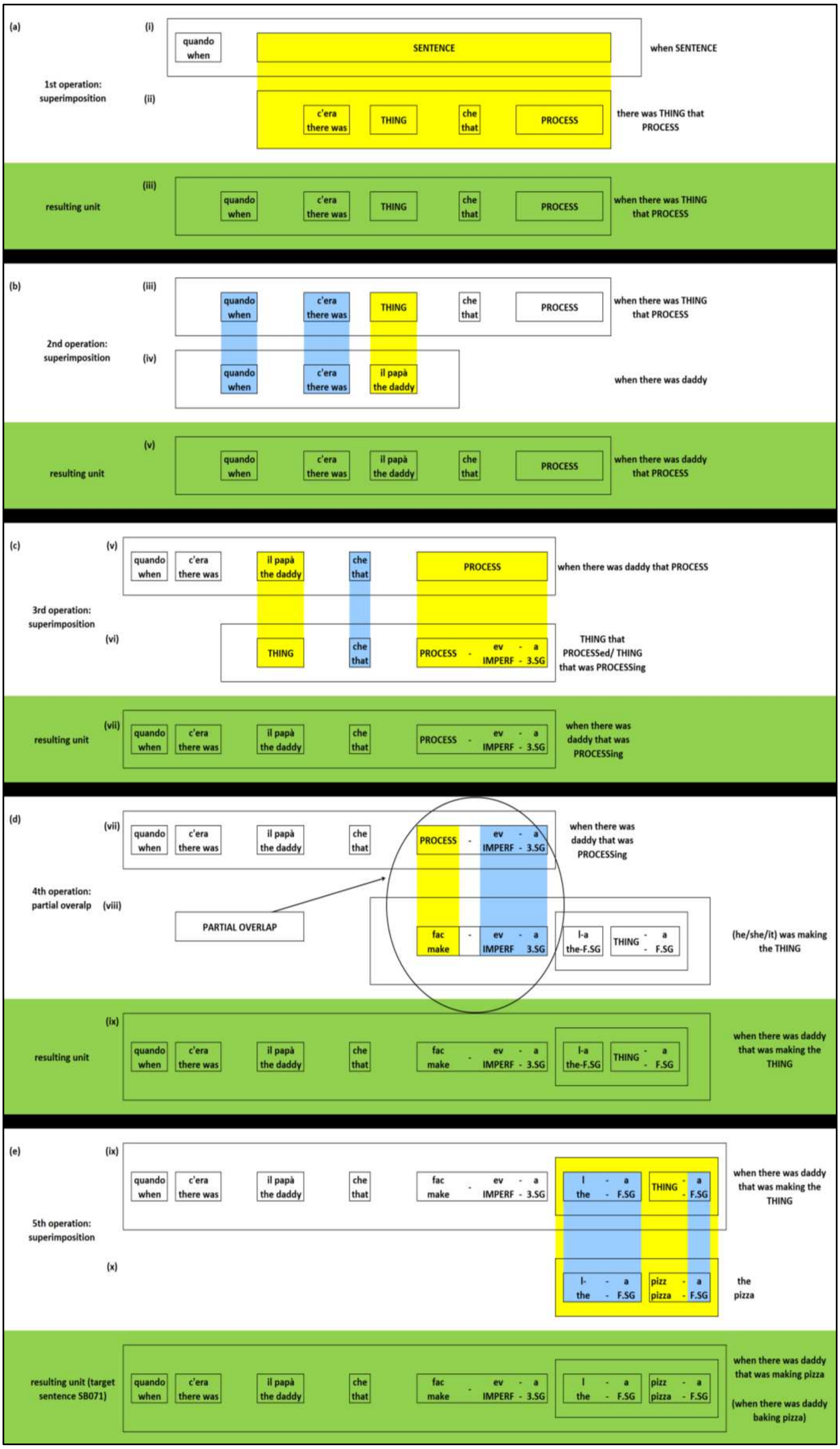


Figure 8.26: deriving target sentence SA071 (69) through partial overlap. Slot elaboration is highlighted in yellow. Shared overlapping material is in blue.

The partial overlaps of attested units can account for twelve sentences that are classified as Hard_Constructional_Fails. In particular, partial overlaps can account for those sentences with an overt subject that are traced back to subjectless schemas. For example, target sentence SB534 (fig. 8.27a) can be traced back to the schema in fig. 8.27b, which is subjectless. Clearly, the pronoun (*noi* "we") cannot be juxtaposed as subject-verb agreement is involved. Hence, the only way to derive SB534 is fig. 8.28.

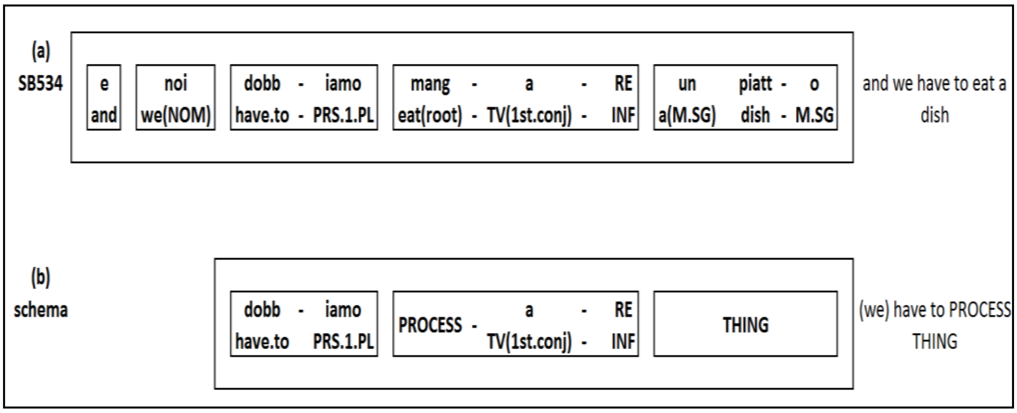


Figure 8.27: target sentence SB534 (a) and the subjectless schema that could account for it (b).

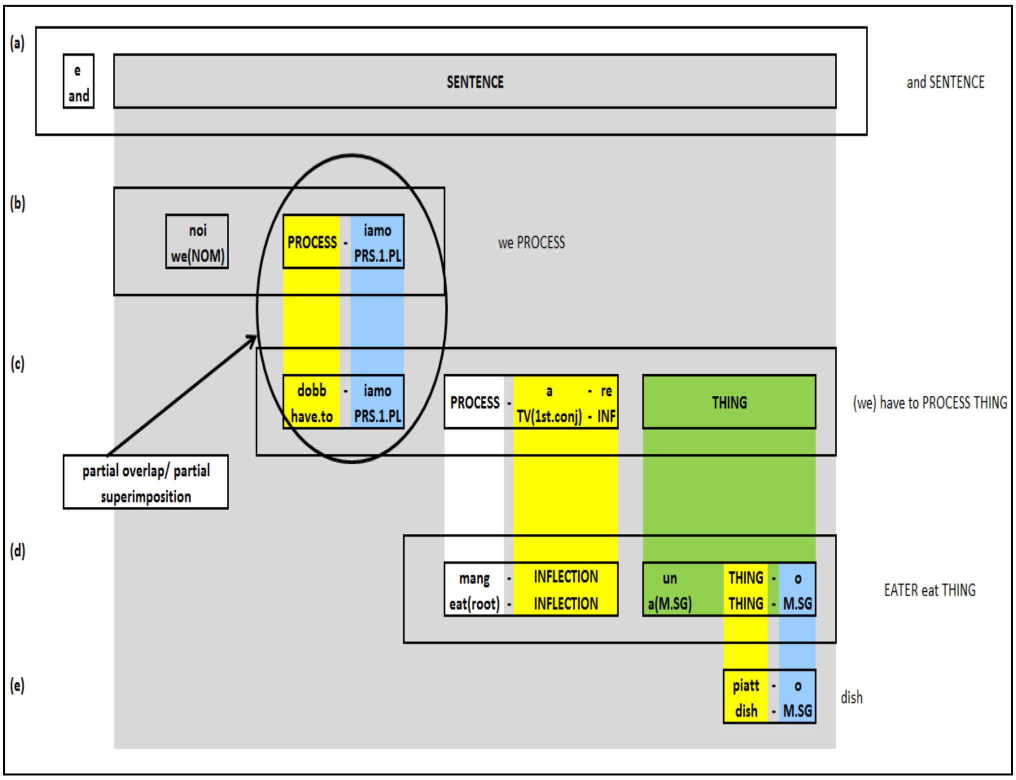


Figure 8.28: deriving target sentence SB534 through partial overlap (in the circle). Elaboration relationships are highlighted in grey, white, yellow and green. Shared concrete material is highlighted in blue.

8.4.5.1. The psychological plausibility of partial overlaps

Unit 8.28b and unit 8.28c are superimposed by virtue of both semantic and concrete correspondences[65], as outlined below:

a) they both instantiate a PROCESS (semantic correspondence);

b) the agentive entity of both schemas is a plurality of *THINGS* of which the speaker is a part (i.e. 1.PL) (semantic correspondence);

c) they share the concrete inflection *–iamo* "-PRS.1.PL", which maps onto the agentive entity of the above point b (concrete correspondence).

The partial overlap of the two units (fig. 8.29a) results in a schema that includes an overt subject acting upon some other entity (8.29b).

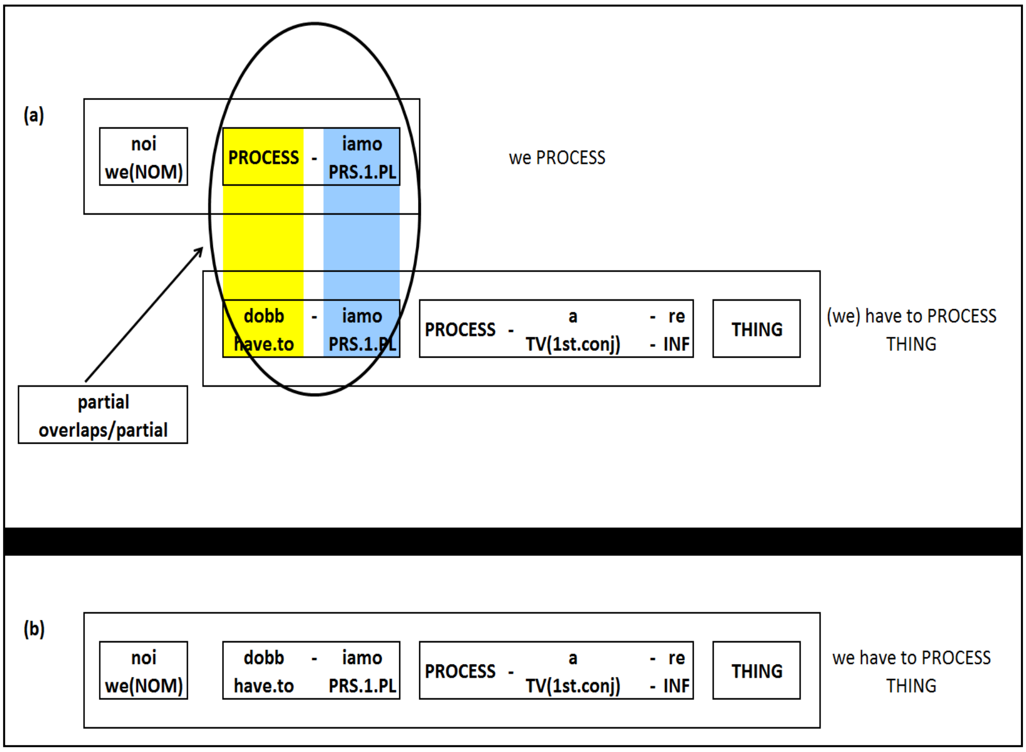


Figure 8.29: the partial overlap of 8.28b and 8.28c (section a) and the resulting string (section b). Slot elaboration is highlighted in yellow. Shared lexical material is highlighted in blue.

Importantly, the partial overlap in fig. 8.29 is possible because the same agent is conceptualised at the semantic and phonological pole of both 8.28b and 8.28c. Fig. 8.30 and 8.31 use a simplistic (and inaccurate if compared to traditional

[65] In Langacker's (2008) terms.

representations) way of representing the two units[66]. Fig. 8.30 shows that, in unit 8.28b, both the pronoun and the verb inflection at the phonological pole map onto the same semantic entity (their semantic poles are linked by a blue horizontal line that indicates this)[67]. Fig. 8.31 shows that *-iamo* "PRS.1.PL" maps onto the same agent as both *noi* "we" and *'-iamo* "PRS.1.PL" in fig. 8.30. Hence, the agent at the semantic pole of the two units is exactly the same and it is also expressed by the same morphological marker (/iamo/). This creates the semantic and morpho-syntactic bases for the partial overlap of the two units and for the appropriate adding of the pronoun *noi* (we) to 8.28c. Overall, partial overlaps account for all those sentences with an overt subject whose putative precedents are subjectless schemas.

Different languages grammaticalise concepts differently by mapping specific meanings, roles and functions onto specific forms. Acquiring a language is about learning a repertoire of units (form-meaning pairings) and **the way they can be "assembled"** (Langacker, 2010; Tomasello, 2003). Consequently, the units speakers learn will differ from language to language and different languages will be acquired differently. It is therefore possible that superimposition could take on different nuances in highly inflected languages such as Italian.

[66] Nonetheless, I believe it is enough for the point being made here. Figures are roughly based on Langacker's (1987, 2008) illustrations.
[67] That is, they are redundant in Langacker's (2008) terms. Redundancy brought about by grammatical elements (e.g. person-number markers) is how a cognitive approach to language accounts for agreement (see Langacker (2008; ch. 6) for an account of it).

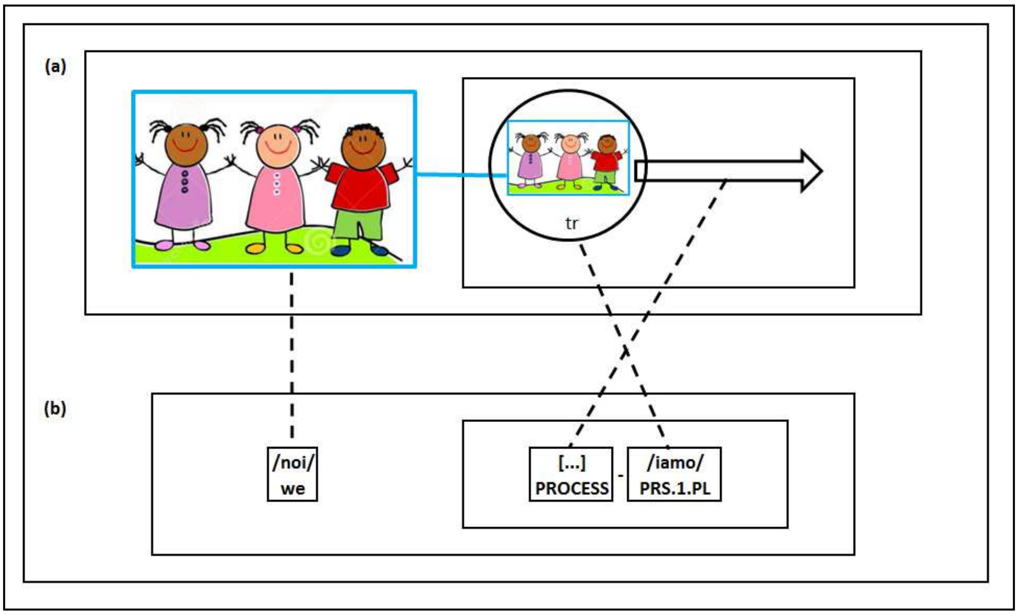


Figure 8.30: the semantic (a) and phonological pole (b) of the unit in fig. 8.28b. The dashed lines indicate relationships of symbolisation. At the semantic pole of both constituents (pronoun and verb), the agent (1.PL) is enclosed in a blue box. The two are linked by a blue line; this indicates that the pronoun *noi* "we" and the person-number marker *–iamo* map onto the same meaning/entity. Translations into English are provided under phonetic transcriptions. [...] indicates that the unit is phonologically unspecified.

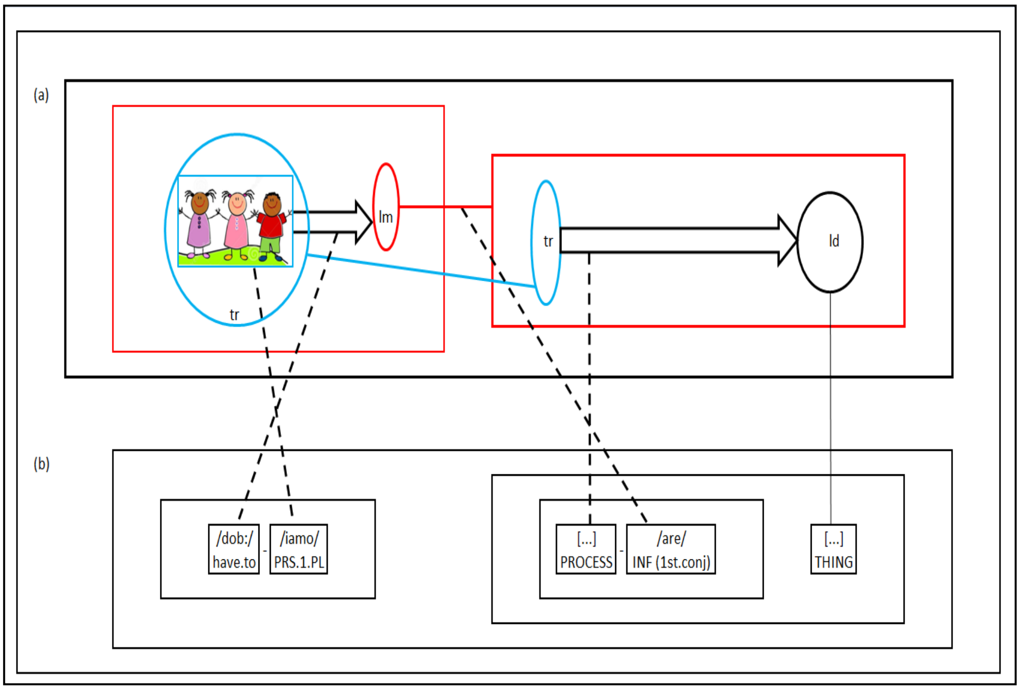


Figure 8.31: semantic (a) and phonological pole (b) of the unit in fig. 8.28c. Dashed lines indicate relationships of symbolisation. The red line that links the landmark of the auxiliary verb (in the red circle) and the dependent *PROCESS* (in the red rectangle) indicates that the infinitive clause is the landmark of the auxiliary. The infinitive marker *–are* is linked by symbolisation (dashed line) to the red line; this indicates that the infinitive marker maps onto the fact that the infinitive clause depends on the auxiliary. The blue box, line and circles indicate that the trajector (agent) of the auxiliary corresponds to the trajector of the dependent clause. Translations into English are provided under phonetic transcriptions. [...] indicates that the unit is phonologically unspecified.

An important aspect to consider when it comes to LA is that successful form-function mapping will also depend on children's cognitive development. In order to attend to specific cues (and hence be able to map forms onto meanings), children must have developed the necessary cognitive abilities on the basis of which they can both attend to those cues and understand the meanings onto which specific forms map (what Bates & MacWhinney (1987) call **functional readiness**).

D'Amico and Devescovi (1993) exposed Italian-speaking children (aged 3;6, 4;6, 5;6, 7;6 and 9;6) and adults to sequences of two nouns and one inflected verb and then asked them to act-out who would act upon whom with toys and props. The design was a 3 (nouns' animacy: AA, AI, IA) x 3 (word order: NVN, VNN, NNV) x 3 (agreement: first noun agrees with the verb, second noun agrees with the verb, both nouns agree with the verb) x 6 (age-group). In Italian, subject-verb agreement is the most important cue to distinguish thematic roles and indeed adults overwhelmingly chose as agent the noun that agreed with the verb (83% of total variance explained by agreement). Agreement accounted for much less variance in 3;6 and 4;6-year-olds (about 4%), who overwhelmingly relied on animacy (59%-87% of total variance explained). By the age of 5;6, children started developing more adultlike strategies (21% of total variance explained by agreement), but it was only at the age of 7;6 and 9;6 that children showed adultlike strategies (75%-88% of variance explained by agreement) [68]. Devescovi and D'Amico's (1993) results were replicated by Devescovi, D'Amico, Smith, Mimica and Bates (1998, experiment 2) who (Devescovi et al., 1998; experiment 3) also tested children with acting-out tasks that exposed them to single-word verbal inflections (e.g. *mang-i* "eat-PRS.2.SG)". From three years of age, children constantly (90%) acted out the meaning of the verb correctly. Hence, 3;6 and 4;6 year-olds understood what verb inflections mapped onto and yet they were unable to use such knowledge in D'Amico and Devescovi's (1993) and Devescovi et al.'s (1998; study 2) experimental tasks.

Devescovi et al. (1998) interpreted such contrasting results in terms of cue cost. When children can attend to local cues (one-word stimuli), they act out single-

[68] Such a late reliance on agreement is attested even in Spanish-speaking children (Kail & Charvillat, 1988; reported in D'Amico & Devescovi, 1993).

word stimuli correctly. When the task requires that children attend to a global cue (subject-verb agreement), they simply ignore it and follow a more local one (namely, animacy).

As Abbot-Smith and Serratrice (2013) note, whether a cue is local or global is a matter of degree. Case-marking is more local than word order, which in turn is more local than subject-verb agreement. In Italian, subject-verb agreement requires that children keep in memory both nouns (subject and DO), their forms (singular vs. plural), as well as the verbal inflection. Subsequently, they have to establish which noun agrees with the verb (the agent) and which does not (the patient) and assign thematic roles accordingly.

D'Amico and Devescovi (1993, p. 288) report that, around the age of six, children move from a local to a global form of control in problem-solving and manual-visual coordination tasks. Verb-subject agreement requires a more elaborated form of global control, as all phrasal elements (and their forms) must be taken into account. Since they have not developed more global cognitive strategies of control in non-linguistic domains, younger children cannot exploit them to attend to a global cue such as subject-verb agreement.

Yet, children as young as two years of age are known to make very few agreement mistakes, even though they have yet to develop more global strategies of control that appear to be necessary to (fully) attend to agreement. It would appear that such an apparent incongruence could be explained by concatenations of partial overlaps.

Concatenations of partial overlaps might be thought of as strategies to conceptualise and apprehend agreement on a more local basis. The partial overlap of fig. 8.28b and fig. 8.28c (depicted in fig. 8.29a) does not require that the child take into consideration all elements of both units (be they schematic or specific). All children have to pay attention to is the shared element *PROCESS-iamo* "PROCESS-PRS.1.PL". Consequently, agreement does not need to be checked against the object slot and against the pronoun (subject). This potentially provides "an escape" from having to factor in all elements involved (subject, verb and object). Correct subject-verb agreement can therefore be delivered by considering only one element of the two units (*PROCESS-iamo*) and is further facilitated by the person-number marker they share (*-iamo*).

Furthermore, when partial overlaps are chosen over full-overlaps (full superimpositions) and juxtapositions, for all target sentences:

a) less operations are needed to trace sentences back to their putative units.
b) there is no need for juxtaposition to be used, as virtually all cases of juxtaposition can be accounted for by partial overlaps (hence the method potentially needs only one type of operation).

Overall then, partial overlaps seem to be child-friendly, as they lighten the cognitive on-line processing load by requiring fewer operations and by allowing the child to deliver correct agreement on a more local basis.

8.4.6. INSERTION

There is another small set of target sentences that could be linked to units attested twice or more, but that could not be derived because operations other than (full) superimposition and juxtaposition would have been needed. Five target sentences seem to have been derived by inserting a component unit into another component unit, which is not permissible under the adopted method (Method_A). For example, target sentence SB128 (70) can be derived by assembling through superimposition the string in fig. 8.32a and by inserting *quasi* "almost" (fig. 8.32b) within it (fig. 8.32).

70. *CHI: *stavo quasi cadendo nel buco.*
    st-av-o *quasi* *cad-e-ndo*
    stay-IMPERF-1.SG almost fall-TV(conj.II)-ing
    ne=l *buc-o*
    in=the(M.SG) hole-M.SG
    "I almost fell into the hole."

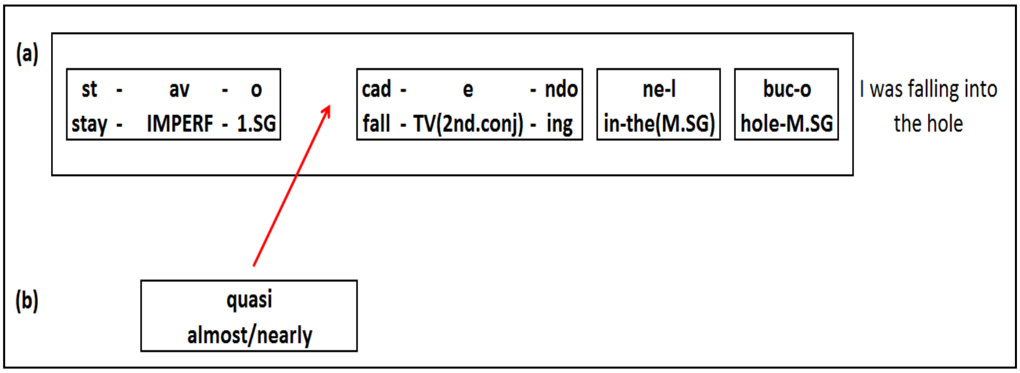


Figure 8.32: deriving target sentences SB128 through insertion.

Even though insertion was used by Lieven et al (2003) as one of five possible operations, positing insertion presents two main problems.

Firstly, how does the child constrain the position the inserted unit_B takes within the unit_A which it "infiltrates"? After all, even a language like Italian presents some constraints on WO (for example the position of clitics, or the order *DET-NOUN*)[69].

Secondly, in order to allow insertion, fig. 8.32a has to be broken down (fig. 8.33a) so that *quasi* "almost" can be inserted (fig. 8.33b). Subsequently, the part of fig. 8.32a that has been set apart must be "glued" back (fig. 8.33c). Hence, insertion brings about the need of breaking and reassembling a unit. This would increase the number of possible operations and would yield a method which would be too unconstrained[70].

Hence, while positing partial overlaps seems to be a realistic and psychologically plausible way of accounting for (Italian-speaking) children's sentences, insertion seems to be too unconstrained, as it is not clear how children would know where to insert a particular unit.

[69] As Dąbrowska (2014) points out, the internal orders of some packets are only **partially** specified.

[70] Indeed, Dąbrowska and Lieven (2005, p. 439) criticised the methodology adopted by that study because it permitted too many operations, which basically allowed Lieven et al. (2003) to derive any sentence through the interaction of many operations (drop, insertion, add on, substitute and rearrange).

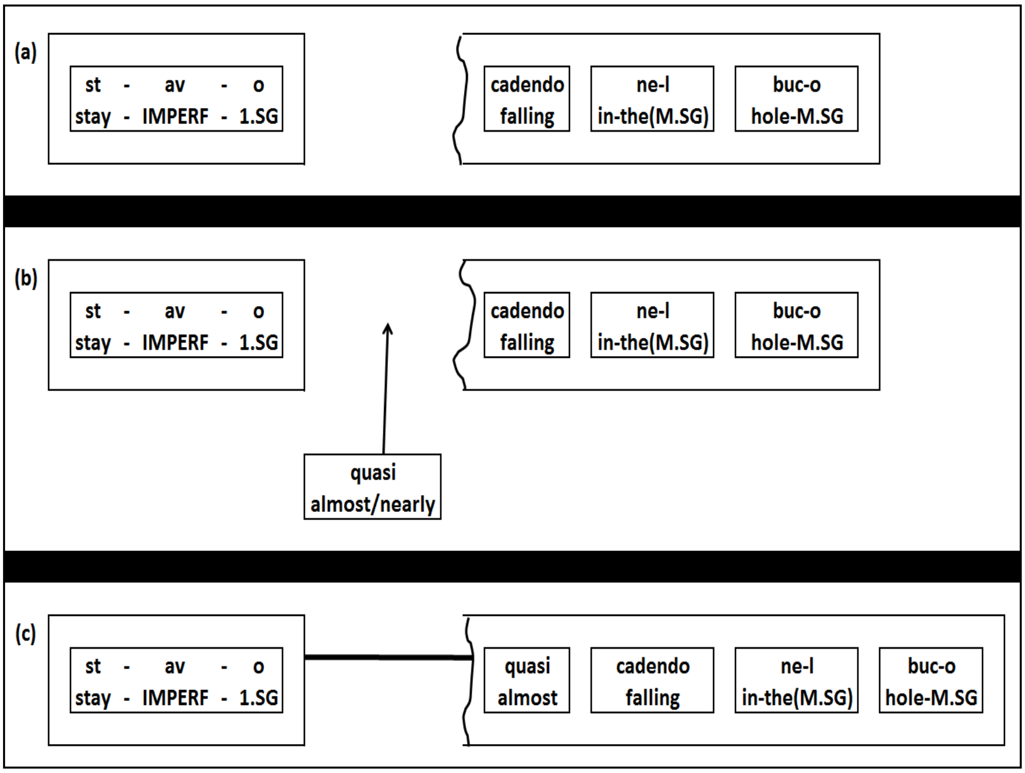


Figure 8.33: the operation of insertion (break, a; insert, b; and re-arrange, c).

8.4.7. PROBLEMATIC CONSTRUCTIONAL FAILS THAT PRESENT MULTIPLE ISSUES

There are about ten sentences that cannot be explained either as extensions of previously encountered strings or by concatenations of attested units. Each of these sentences presents a combination of the issues discussed thus far. For example, they could be derived by inserting a unit into a putative schema whose slot did not meet the variance requirements; or by a concatenation of schemas attested only once. Six of these are grammatical and four are not. If summed up with cases of insertion, there are fifteen sentences (ten grammatical and five ungrammatical) that are problematic under many respects and they represent 2% of the whole dataset.

As previously pointed out (6.4.2.1), the sampling regime of this study captures 5% to 10% of a child's linguistic experience. Hence, it might be the case that the study failed to capture on tape fifteen construction types used by the child, in the same way as it failed to capture twenty-seven lexical items (lexical fails under Method_D; table 20.1 and fig. 20.1 in Appendix_II), the latter being necessarily learnt from the child's own previous linguistic experience.

8.5. CONSTRUCTIONAL FAILS AND SYNTACTIC COMPLEXITY

The analysis thus far undertaken has shown that virtually all syntactic fails can be accounted for in terms of narrow-scope and semantic, rather than syntactic, generalisations. It is nevertheless also true that those sentences could not be fully accounted for by using Method_A. That is, they could not be fully explained in terms of lexically-specific units. Hence, one might argue against the qualitative analyses put forward in sections 8.1-8.4 by claiming that such sentences are evidence that Roberto relies on fully-schematic templates (possibly even innate principles and constraints) that cannot be fully characterised in lexically-specific terms.

It is therefore reasonable to ask whether there is any evidence of reliance on more schematic (abstract) templates in those constructional fails. One way to enquire into such an issue is to investigate whether constructional fails are structurally more complex than successful derivations. The rationale for choosing syntactic complexity as a measure of the degree of schematicity on which the child relies is that, if constructional fails presented more complex structures, one might argue that such a more structural complexity is evidence that the child is capable of going beyond the lexically-specific strings he encounters by drawing more complex and more adultlike generalisations from them.

In order to measure syntactic complexity in successful derivations and constructional fails, the two types of target sentences were compared by analysing the structural complexity of the t-units (Hunt, 1965) they contained.

8.5.1. USING T-UNITS TO MEASURE SYNTACTIC COMPLEXITY

Following Hunt (1965), a t-unit is defined as a main clause and any non-clausal structures (e.g. PPs) and/or subordinate clauses that are either attached to or embedded in the main clause (see Appendix_II, ch. 20.3.3, for examples of t-units).

Hunt (1965) investigated the development of syntactic complexity in the writing of fourth, eighth and twelfth grade English-speaking children using various measures of complexity (e.g. length in words of sentences, number of clauses per sentence and length in words of t-units). He found that the best index of syntactic growth was the mean length in words of t-units. Nippold, Hesketh, Duthie and Mnasfield (2005) further found that such a measure is a good predictor of the use of subordination; for instance, the mean length in words of t-units was positively correlated with the use of both relative ($r=.46$ $p<.05$) and adverbial (r.=59, $p<.01$) clauses in the linguistic production of eight-year-olds.

Since the complexity of t-unit is often used as a (reliable) measure of syntactic development, it is here adopted to analyse the syntactic complexity of successful derivations and constructional fails.

8.5.2. METHOD

Firstly, each t-unit **type** was identified. A t-unit was defined in both semantic and grammatical terms. Semantically, it had a meaning that was complete and independent; that is, t-units had to have a coherent semantics so that their meaning could be (mostly) inferred without the support of other sentences. Grammatically, a t-unit had a main clause (with its attached and/or embedded subordinate clauses and non-clausal structures) that presented at least one verb inflected for person (1st, 2nd, 3rd) and number (plural vs. singular). Appendix_II (ch. 20.3.3) provides the reader with an accurate report of how t-units were identified and the criteria for considering a target sentence as presenting one, two, or more t-units.

Overall, 113 sentences that either did not present a verb inflected for person and number or were verbless were excluded from the analysis as they did not contain any t-unit. A further 31 sentences were classified as fails because they contained a pure lexical fail and were therefore excluded as the purpose of the analysis was to compare successful derivations and constructional fails.

This left 624 target sentences with a total of 646 t-unit types, 90 of which were contained in constructional fails, whereas the remaining 556 were contained in successful derivations.

Once the t-units to be analysed were identified, two measures of syntactic complexity were calculated: number of words per t-unit and number of (finite) clauses per t-unit. Hunt (1965) suggests that the proportion of subordination in a sample can be calculated by subtracting 1 from the ratio of clauses per t-unit:

[1- (number of t-units / number of clauses)]. Hence, a ratio of 1.03 means that 3% of t-units present at least one subordinate clause.

Loosely following Hunt's (1965) methodology, a clause (either main or subordinate) was defined as a structure that contained a verb marked for person and number. Structures that contained infinitive verbs were not considered as clauses. Hence, (71) is a single t-unit containing a main clause (*sai che* "do you know that") and three subordinate clauses. Conversely, (72) is a t-unit containing only one clause as its subordinate clause (*per entrare nella mia casetta* "to enter into my little house") is an infinitive clause (see Appendix_II, ch. 20.33).

71. **CHI: < sai> [/] sai che io (.) ho fatto una corsa (..) grande + e sono arrivato (.) dal Luca +. +. e ho fatto pf@o.*
sai che io ho fatto una corsa grande e sono arrivato dal Luca e ho fatto puf.

sa-i	*che*	*io*	*ho*	*fatto*	*un-a*
know-PRS.2.SG	that	I	have(PRS.1.SG)	done	a-F.SG

cors-a	*grand-e*	*e*
run-F.SG	big-SG	and

sono	*arriv-a-t-o*	*da=l*
be(PRS.1.SG)	arrive-TV(conj.I)-PTCP-M.SG	at=the(M.SG)

Luca	*e*	*ho*	*fatto*	*puff.*
Luca	and	have(PRS.1.SG)	done	puff.

("You know that I have done a big run and I have arrived at the Luca and I've done ʽpuff'?").
"Did you know that I ran a long way and got to Luca and went ʽpuff'?"
(week5.2014.02.18.B: lines 1423 and 1425)

72. **CHI: No però sei troppo piccola per entrare nella mia casetta*

no,	*però*	*sei*	*troppo*	*piccol-a*	*per*
no ,	but	be(PRS.2.SG)	too	small-F.SG	to

entr-a-re	*ne=ll-a*	*mi-a*
enter-TV(conj.I)-INF	in=the-F.SG	my-F.SG

cas-ett-a
house-little(ENDR)-F.SG
"No, but you are too small to enter into my little house."
(week5.2014.2.18.B: lines 2334-2335)

8.5.3. RESULTS AND DISCUSSION

Since neither words per t-unit nor clauses per t-unit were normally distributed in either successful derivations or constructional fails (Shapiro Wilk, W=0.14 to 0.91, p<.05), the analysis was conducted with one-tailed independent Wilcoxon rank sum tests. As table 8.5 shows, constructional fails present more complex t-units than successful derivations under both measures of complexity.

Table 8.5: comparing the complexity of the t-units identified in successful derivations and constructional fails.

T-UNITS COMPLEXITY: comparing successful derivations and constructional fails			
measure of complexity	successful derivations mean (SE)	constructional fails mean (SE)	Wilcoxon rank sum test with continuity correction (one-tailed)
number of words per t-unit	4.29 (0.73)	*6.15 (0.23)*	*W=38***
number of finite clauses per t-unit	1.03 (0.01)	1.11 (0.04)	W=27**

** p<.05

One may argue that the greater complexity of Constructional_Fails with respect to successful derivations seems to indicate that Roberto is able to infer abstract (possibly syntactic and adultlike) generalisations from the strings he encounters. Consequently, it could be argued that he uses those generalisations to produce more complex structures that cannot be (fully) characterised (explained) in strictly lexically-specific terms[71].

Although such an interpretation may be considered legitimate, I believe it would fail to address the larger picture that emerges from the analyses undertaken thus far. For instance, an important aspect to bear in mind is that both Soft and Hard Constructional_Fails present a higher proportion of ungrammatical, non-adultlike sentences than successful derivations (table 8.4). Moreover, a more qualitative insight into constructional fails appears to suggest that such ungrammatical sentences seem to be underpinned by the same types of narrow-scope and semantic generalisations that characterise successful derivations.

Superficially then, there appear to be two contrasting results. On the one hand, both the grammaticality analysis (table 8.4) and a more qualitative analysis of Constructional Fails (refer back to 8.4.3) suggest that Roberto has yet to develop adultlike, syntactic generalisations. On the other, the fact that sentences that cannot be traced back contain more complex t-units than successful derivations (table 8.5) might be regarded as evidence that Roberto is capable of generalising

[71] Note that such an interpretation is only possible because those more complex sentences cannot be fully accounted for in lexically-specific terms. Indeed, knowledge of very complex structures (e.g. long-distance questions) may indeed be characterised in terms of lexically-bound templates (see Dąbrowska, 2004).

beyond the lexically-specific knowledge that could be inferred from his language input.

However, I would argue that such results are not inconsistent with each other and the fact that constructional fails are syntactically more complex than successful derivations does not represent conclusive evidence that the child is relying on abstract (possibly adultlike) syntactic (rather than semantic and functional) generalisations.

The greater structural complexity of constructional fails could indeed be interpreted as evidence of Roberto attempting to go beyond strictly lexically-specific knowledge. However, the key factor is the nature of the generalisations on the basis which he does so and the extent to which such generalisations yield adultlike, well-formed sentences.

It seems to be the case that, when pressed by communicative needs that appear to require more complex structures than the ones he seems to have mastered, Roberto does indeed try to go beyond the lexically-specific generalisations he has learnt. However, it also appears to be the case that he does so by relying on form-function generalisations that are still semantic, rather than syntactic. For instance, he appears to assimilate a pronoun (*qualcun-a* “someone-F.SG”) to an adjective (*qualche* “some”) on the basis of phonological (/kwualk/) and semantic (indefinitiveness) similarities. The result is that when he attempts to produce more complex sentences, he is more likely to utter ungrammatical sentences. Thus, as the elaborative distance between TS and SS increases, so does the likelihood of delivering ungrammatical sentences. This appears to suggest that Roberto has yet to develop those types of more mature (adultlike) generalisations that would allow him to deliver more complex sentences in a way that is consistently adultlike.

Page intentionally left blank

9.

DISCUSSION AND CONCLUSION

9.1. SUMMARY AND DISCUSSION

9.1.1. RESEARCH QUESTION

a) Can Italian-speaking children's early language be accounted for in terms of lexically-specific units acquired from the concrete language that children themselves have previously experienced?

To the extent that Roberto could be said to be representative of two-year-old Italian-speaking children, 82% of what an Italian-speaking child says can indeed be accounted for in terms of lexically-bound units. This leaves out a significant part (18%) of children's speech that is not straightforwardly explained by the method. About 13% of Roberto's target sentences are classified as constructional fails; the method failed to find appropriate constructional derivations for ninety-seven sentences. Interestingly, constructional fails contain more complex t-units than successful derivations. A possible interpretation of such failures is that those sentences have been generated by drawing on an abstract and innate language-specific device.

Although this may be possible, I agree with Langacker (2000, p. 2) who claims that an innate language faculty should be invoked only as last resort. A more careful analysis of fails seems to indicate that such a last resort is not needed to account for Roberto's target sentences. Indeed, a cognitive approach to language accommodates virtually all of them (98%).

9.1.2. FORM-FUNCTION MAPPING, UNGRAMATICALITY, CREATIVITY AND DYNAMICITY

The Method, which assumes lexically-based generalisations based on form-function mapping, accounts for 87% of the syntactic patterns used by Roberto. It also accounts for his morpho-syntactic overgeneralisations (see 8.1). Schemas are productive units on which children rely; as such they might lead to cases of over-productivity. Overall though, they allow children to infer the right kinds of generalisations and therefore produce mostly (98%, fig. 8.15) adultlike grammatical sentences (see Dąbrowska, 2000; Lieven & Brandt, 2011; Lieven & Tomasello, 2008, for similar observations). Interestingly, both overgeneralisations attested seem to be instantiations of the **caused-motion construction** (Goldberg, 1999, 2006). Such generalisations, that seem to stem from semantics-based analogies, *do* however interact with Roberto's lexically-specific units, from which both sentences can be derived. Goldberg et al. (2004) showed that Argument-Structure-Constructions are dominated by specific light verbs, which represent the prototypes from which such constructional patterns inherit their meanings. Such a hypothesis seems to be supported by Roberto's overgeneralisations and some schemas yielded by the method. Both overgeneralisations – (51) and (52) – are instantiations of the caused-motion-construction, whose prototype verb is *put* (fig. 8.5). Furthermore, sentence (51) (**descend it down* or **go it down*) can be thought of as an extension vis-à-vis a prototypical instantiations of the cause-motion-construction (*put it down*) (fig. 8.4). Finally, four schemas built around the verb *mettere* "to put" that instantiate the caused-motion-construction have been retrieved in the test corpus (Appendix_II, fig. 20.17). Hence, both schemas built around *mettere* "to put" and Roberto's overgeneralisations present the same argument structure.

Goldberg (1999) hypothesises that the development of such kinds of schematisations is likely to be a by-product of vocabulary growth.

> It is likely that the categorization and generalization into more abstract patterns is driven by an increase in vocabulary size. That is, in order to learn an ever increasing vocabulary and the associated syntactic patterns, it may be necessary to categorize individual instances into classes.
>
> Goldberg (1999, p. 206)

As previously discussed, Caselli et al. (1999) and Devescovi et al. (2005) found evidence that grammatical development and vocabulary growth go hand in hand. The fact that Roberto is found to have an above average vocabulary is consistent with Goldberg's hypothesis. Thus, his large vocabulary size might be bringing about the need to start organising his inventory in a more interconnected (schematic) network.

Roberto's large vocabulary and the fact that Italian presents a highly regular and frequent nominal morphology account for the fact that the child could be said to be productive with this aspect of the morphological system. The nominal morphology of his ambient language represents a local cue (low processing cost), which is highly available and reliable in Bates and MacWhinney's (1987) terms. The interaction of those characteristics predicts that regular morphology should be acquired quite early on in development and indeed all morphological fails are regular nouns.

Roberto's generalisations are nevertheless still functional and semantic, rather than syntactic in nature, and do not appear to be rooted in an adultlike ability to parse the input. A case in point is represented by the two sentences analysed in section 8.4.3. In target sentence SB468 (59), Roberto uses the pronoun *qualcun-a* "someone-F.SG" instead of the adjective *qualche* "some". The hypothesis that has been put forward is that the attested unit *qualche THING-a* "some THING-f.sg" is the prototype - in Goldberg's (1995, 1999, 2006) terms - from which the construction *X THING* is inferred. Because of both semantic (indefinitiveness) and formal (the string of phonemes /'kwualk/) similarities, the distributional patterns of the adjective are extended to the pronoun. One might speculate that in extending the distributional patterns of *qualche* (pron.) to *qualcuna* (adj.), Roberto is (tentatively) inferring (and using) a more complex (schematic) construction (*X THING*) on the basis of a form-function mapping that is still not adultlike. The result is a string that is not grammatical.

Such an interpretation is consistent with two other findings that emerge from the analyses just presented.

Firstly, Constructional_Fails contain longer (more words) and more complex (more clauses) t-units than successful derivations (table 8.5). In a sense, this result could be regarded as evidence that Roberto is attempting to go beyond the lexically-specific patterns he experienced (possibly pushed to do so by his large vocabulary).

Secondly, both Soft and Hard Constructional_Fails show higher rates of ungrammatical sentences than successful derivations (fig. 8.15; table 8.4).

This suggests that when Roberto attempts to go beyond the lexically-specific patterns he experienced, he does so by relying on generalisations that might be the wrong ones and hence likely to yield ungrammatical sentences. Such an interpretation appears to be supported by the analysis of those sentences whose agreement between subject and other phrasal constituents (participles and adjectives) cannot be (fully) accounted for using Method_A (section 8.4.4.2). When sentences that are classified as Soft_Constructional_Fails (which therefore have putative precedents that could account for agreement in lexically-specific terms) are taken out of the equation, the remaining subgroup of sentences contains more wrong (60%) than correct (40%) agreements. This suggests that Roberto mastery of agreement is still lexically-bound, at least to some extent.

Overall, the above interpretations appear to lead to two main general observations:

a) Roberto is still unable to parse the input according to adultlike syntactic categories. Were he able to do so, he would not have assimilated a pronoun to an adjective.

b) children *are* creative language learners. When they are creative however, they might struggle to infer the right kinds of generalisations, becoming increasingly more likely to utter ungrammatical sentences.

A further observation is that the dynamicity of the system must always be taken into account. Status of units is a by-product of cognitive salience and entrenchment. Entrenchment is a continuous, on-going cognitive process that changes as a function of frequency. Hence the representational strength of a unit (its entrenchment) is a matter of degree. Following research in non-linguistic domains (e.g. Hebb, 1961), Langacker (2010) argues that, as a processing activity (the retrieval of a unit to categorise a new target) is undertaken more and more frequently, its representational strength increases, and engaging in such an activity becomes more and more likely and easier. This is consistent with Munakata et al.'s (1997) characterisation of knowledge in terms of graded representational strength (refer back to 3.10.1). Put simply, the more a unit is evoked to categorise new expressions, the more likely and easier it will be to evoke it again to categorise further new expressions.

A hypothesis that may be said to be consistent with such a dynamic characterisation of units' entrenchment is that linguistic units that are more entrenched (more frequent) should be easier to activate and, because of that, they should also be more likely to yield adultlike, grammatical sentences. Since those units are instantiated by more strings, they are likely to represent more solid generalisations. Conversely, units that are less entrenched should be more difficult to retrieve and hence more likely to yield non-adultlike targets. Assuming that the frequency in the main corpus of the constructions identified somehow mirrors their frequency in the actual child's experience, it is possible to hypothesise a (continuous) rank of representational strength, i.e. entrenchment, i.e. ease of activation:

73. successful derivations > Soft_Constructional_Fails > Hard_Constructional_Fails.

Consequently, one could speculate that successful derivations, Soft_Constructional_Fails and Hard_Constructional_Fails represent a continuum of entrenchment (representational strength), which is mirrored (or yields) a continuum of adultlike (grammatical) targets (98%, 85% and 66% of grammatical sentences, respectively). This hypothesis is supported by the statistical analysis, which showed that both types of Constructional Fails are more likely to present ungrammatical sentences than successful derivations ($p<$ .05; table 8.4).

Importantly, when the "not-so-entrenched" model is not too distant from its target, the extension process might well lead to well-formed (grammatical and adultlike) sentences – e.g. (54). Conversely, when the elaborative distance is greater, as in the cases of (58) and (59), the generalisations inferred might not be the right ones. The results are ungrammatical and non-adultlike sentences.

9.1.3. IS THE METHOD PSYCHOLOGICALLY PLAUSIBLE?

Clearly, a quantitative method that is based on a dichotomous distinction "fail vs. successful" has limits. Whether such a method is in fact a psychologically realistic way of accounting for a cognitive activity such as language is an open question to be addressed by other research. However, when a more qualitative insight into constructional fails is provided, a cognitive and UB approach to language can handle those sentences quite well in light of Roberto's previous linguistic experience. The exception being fifteen sentences (2% of the dataset) that are problematic to account for. As for cases of insertion however, it has to be pointed out that those target sentences *could* be accounted for by lexically-specific patterns available in the input; they are unsuccessful derivations “simply” because they could not be assembled by applying superimposition and/or juxtaposition

Lieven et al. (2003), whose method included insertion, found that such an operation was only marginally needed to trace the child’s sentences back to units in the main corpus. However, when the mother’s corpus was analysed, the use of insertion was needed more often. They suggest (Lieven et al., 2003, p. 30) that insertion might reflect a more advanced linguistic ability and, possibly, some more mature grasp of constituency. The details of how such a more mature grasp of constituency is acquired is another open question.

Partial overlap is also an operation that has not been allowed under the adopted method (Method_A). Yet, unlike insertion, it seems to be a realistic candidate to analyse the way in which children could assemble their sentences, particularly when a highly inflected language such as Italian is at issue. Partial overlaps allow agreement to be explained on a local basis, and local strategies of problem solving seem to be more attuned to young children’s cognitive abilities. Hence, they provide a coherent picture as to why children make so few agreement mistakes in their spontaneous speech and yet they struggle to use such knowledge while attending some experimental tasks (e.g. D’Amico & Devescovi, 1993; Devescovi et al., 1998, experiment 2).

Different languages are acquired differently, and the linguistic representation speakers of one language might have is likely to be different from the representation developed by speakers of another language. Hence, certain operations posited by UB researchers (such as superimposition) might well be exploited differently, depending on the language at stake (hence superimposition in Italian might well be based on both full and partial overlaps).

9.2. CONCLUSION

Overall, the results of Roberto's traceback seem to support a UBA to LA as the vast majority of sentences (82%) could be derived from lexically-specific units he had previously encountered. These results mirror previous results regarding English-speaking children (Dąbrowska & Lieven, 2005; Lieven et al., 2003, 2009).

Nearly 90% of the syntactic patterns used by the child were *Schemas-with-Slots* or *Fully Lexically-specific Strings* that were classified as available units. This suggests that the input children receive is rich enough to allow them to infer the right kinds of generalisations to become competent speakers. Furthermore, even those cases that were classified as Constructional_Fails could be accounted for by a Cognitive Linguistics Framework and by a dynamic view of entrenchment (i.e. status of unit). Overall, all sentences but fifteen (2% of the whole dataset) can find a principled explanation within the UBA adopted by this research on the basis of Roberto's own concrete experience. That is, even though some sentences (18%) cannot be derived using Method_A, the overwhelming majority of those unsuccessful derivations can nonetheless be explained in terms of input-based and narrow-scope form-function generalisations. Thus, it appears to be the case that the child's linguistic production can be accounted for without positing innate, abstract linguistic knowledge.

Clearly, one could still challenge such an interpretation by claiming that these results are by-products of the interaction between children's routinised activities and the statistical and distributional properties that are attested in both children's and adults' language use (see Yang, 2009). Such an argument would be consistent with the fact that Constructional_Fails present more complex t-units than successful derivations (table 8.5). However, this appears to be more difficult to fully maintain when a more qualitative and thorough analysis is undertaken.

Firstly, most Constructional_Fails seem to be underpinned by the same kinds of lexically-based form-function generalisations in which successful derivations are putatively rooted. These generalisations appear to have a functional, rather than syntactic nature. As for Constructional_Fails, some of these generalisations are also based on strings attested only once in the main corpus and hence on supposedly less entrenched units. The interaction between the semantic nature of these generalisations and the (relative and supposed) low frequency of the strings on which they are (putatively) based seems to lead towards ungrammatical and non-adultlike sentences; particularly when the TS seems to present greater elaborative distance from its SS.

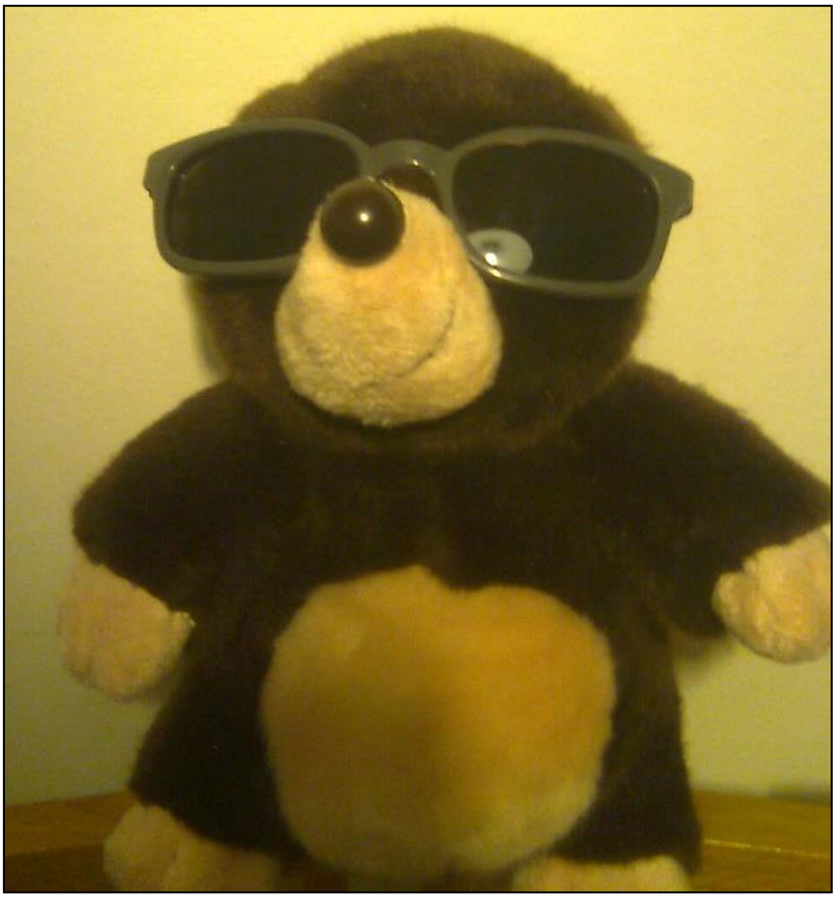
Secondly, a more careful look into Roberto's ungrammatical sentences is also consistent with UB models, whereas they appear to be difficult to explain away in terms of routinised situations and the inescapable distributional properties of language. Indeed, both Soft_Constructional_Fails and Hard_Constructional_Fails present a higher proportion of non-adultlike sentences ($p < 0.05$) than sentences that can be thought of as instantiations of well entrenched schemas (i.e., successful derivations; fig. 8.15). If successful derivations were by-products of the statistical properties of language use and routinised situations, why would those sentences that have less precedents (and/or present greater elaborative distance from their putative SSs) be more ungrammatical than sentences that are successfully accounted for?

Yet, the extent to which the spontaneous production of only one child is representative of the whole population of Italian-speaking children is at least questionable. As discussed in 3.7, spontaneous production sampling can only report part of what children say and hear. It does not provide exhaustive evidence of what children can and cannot say and the kinds of linguistic connections and generalisations of which they are capable.

In order to enquire into such questions, experimental evidence that carefully controls for the kind of input infants receive is needed. Part III provides this very much needed evidence.

Part III

An Experimental Insight into the Development of Morpho-Syntactic Competence



Page intentionally left blank

10.

INTRODUCTION

The results of the naturalistic study are consistent with UBAs to LA, as most of the language spoken by Roberto can be accounted for in terms of low-scope, lexically-specific generalisations (schemas) inferred from the specific input he experienced.

The second part of this research reports on an experimental study that investigated the development of morpho-syntactic competence in Italian-speaking children aged 2;02 to 5;00.

Previous results regarding English-speaking children suggest that they develop adultlike schematic competence gradually. For instance, even though sensitivity to the transitive pattern is attested as early as the onset of the third year of life (Tomasello & Abbot-Smith, 2002; Noble et al., 2011), it is not before another two to three years that children are constantly able to reject ungrammatical orders to favour the grammatical schemas of English (Akhtar, 1999; Abbot-Smith et al., 2001).

The study to be presented investigates whether the acquisition of Italian can be described in terms of a piecemeal development that gradually leads to a fully-schematic (or adultlike) language. In order to do so, the study attempts to enquire into the following research question:

> *To what extent can Italian-speaking children be said to rely on (have mastered) fully-schematic constructions/patterns?*

The above research question has been broken down into two sub-questions:

a) *Is there any evidence that Italian speaking children can rely on (retrieve) fully schematic constructions in order to be* ***morphologically*** *productive?*

b) *Is there any evidence that Italian speaking children can rely on (retrieve) fully schematic constructions in order to be* ***syntactically*** *productive?*

Furthermore, the relationship between vocabulary (linguistic experience), age (maturity) and morpho-syntactic productivity (grammatical development) has been investigated through the research question below:

c) *Given previous findings regarding both Italian and English-speaking children, will vocabulary be a better predictor of morpho-syntactic productivity than age?*

Since investigating children's productivity with all aspects of either syntax or morphology would clearly be too broad a research question, research question (a) is answered by enquiring into the extent to which children can be said to be productive with past participles of Conjugation I verbs (that is, the extent to which they can be said to have acquired the schema *PROCESS-a-to* "PROCESS-TV(conj.I)-PTCP"); and research question (b) is investigated by tapping into the development of the transitive construction. Research question (c) has been investigated through correlation analyses that enquired into the relationship(s) between age, vocabulary and morpho-syntactic productivity.

Before proceeding I must demonstrate the system of phrasal notation that is to be employed throughout the experimental analysis. When reference is made to a Direct Object (DO) with a *CAPITAL LETTER* (VO), this is to indicate a full NP or a strong pronoun. When I refer to a DO with a lower case letter (oV, Vo), the DO is a clitic ACC pronoun. When schemas or patterns are reported, "<" indicates a fixed linear sequence, whereas "+" indicates that the linear order is unspecified. Elements that are included in brackets are optional. Thus, (*S) + V<O* means that the fixed sequence VO can either precede (VOS) or follow (SVO) an optional S.

11.

METHOD

11.1. PARTICIPANTS

The participants were ninety-three children (age 2;02-5;00) and fifteen adults (age 24-43) who were native-speakers of Italian and lived in the north-east of Italy. Data were collected during a preliminary study (eighteen children) and a main study (all remaining participants). One two-year-old child was recruited privately by contacting his/her parents through mutual acquaintances, while all remaining children were recruited through five local nurseries using appropriate legal procedures. Adults were friends (5) or acquaintances (5) of the experimenter, or members of his extended family (5) (see table 21.1 in Appendix_III).

For various reasons, fourty-one participants had to be excluded from the study; twenty-seven children because they were not compliant with the task, three children because they turned out to be native speakers of other languages, five children and one adult because of experimenter error and one child due to his/her unintelligible speech. This yielded a group of sixty-seven participants, divided into four age-groups (table. 11.1)

Table 11.1: the participants who took part in the experiment divided by age group.

age group	gender of participants		total no. of participants	age of participants (years:months:days)	
	males	females		mean	range
two-year-olds	9	7	16	02;08	02;02.23 - 03;00.25
three-year-olds	11	5	16	03;08	03;01.16 - 03;11.13
four-year-olds	9	12	21	04;05	04;00.28 - 05;00.00
adults	5	9	14	29;00	24;10.19 - 43;10.07
total	34	33	67		

11.2. DESIGN AND MATERIAL

11.2.1. DESIGN

11.2.1.1. productivity ~ age*verb_familiarity*construction

As already briefly discussed in chapter 4, the experiment to be presented elicited both morphological and syntactic productivity in Italian-speaking children and adults. Participants were exposed to both a familiar verb and a nonce verb; each verb was combined with a physical transitive action and was presented in its IMP.2.SG form in combination with

either

a) a post-verbal clitic DO ("Vo" construction) - *Verb-a=l-a!* "Verb-IMP.2.SG=clitic.3.ACC-F.SG!" ("Verb her!"),

or

b) an agent that was expressed in the form of an addressee to whom either the child or the Experimenter (E.) ordered to perform the action ("V,A" construction) – *Verb-a, [TOY'S NAME]!* "Verb-IMP.2SG, [TOY'S NAME]!" ("verb, Agent/Addressee!")[72].

Morpho-syntactic productivity was then elicited by prompting participants to use those verbs in past tense (*passato prossimo*) transitive constructions – i.e. *(S)+AUX<V<O* or *(S)+o<AUX<V*. The design was therefore one in which linguistic productivity was measured as a function of:

a) whether the verb was familiar or nonce (verb_familiarity)
b) whether the verb was presented within a "Vo" or a "V,A" construction (construction)
c) the age of the participants (age).

Since syntactic and morphological productivity were coded separately, the final experimental design had either syntactic or morphological productivity as dichotomous DV (productive vs. not productive).

[72] The label "A" in "V,A" indicates the double function of the element. It is the agent of the action and the addressee in the construction.

The predictors were age[73], construction (“Vo” vs. “V,A”; repeated-measure) and verb_familiarity (familiar vs. nonce; repeated-measure):

productivity ~ age*verb_familiarity*construction

11.2.1.2. A brief introductory overview of task and design

All children attended two two-hour familiarisation group sessions and one or two fifteen to thirty-six minute individual sessions. The individual session, called **test phase**, consisted of three parts. In the first part, called **warm-up,** children were introduced to the experimental task and learnt the names of the toys and props with which they would play. In the second part, called **training**, children were familiarised with the task and were exposed to two real, existing verbs (called **training-familiar verbs**). In the third part, which constituted the proper **experimental phase**, children were first exposed to a real, **familiar** verb and then to a **nonce** one. Because the familiar verb acted as a control condition, I shall refer to it as **control-familiar verb**. Similarly, since the nonce verb constituted the test condition, I refer to it as **test-nonce verb**.

Each verb (the two training-familiar verbs, the control-familiar verb and the test-nonce verb) was combined with a transitive action performed by either “Emily Elefante” (in English “Emily Elephant”) or “Peppa Pig”. When one character was assigned agent role, the other acted as patient. Each verb-action pair was presented within either a “Vo” or a “V,A” **construction**. The assignment of each action-verb pair to either “Vo” or “V,A” construction, the character who was assigned agent role (*Peppa Pig* versus *Emily Elefante*) and the training verbs’ order of presentation were counterbalanced across participants in each age group. For each participant, the control-familiar verb and the test-nonce verb always appeared in a different construction (either “Vo” or “V,A”) and had a different agent, as did the two training-familiar verbs.

[73] Age was coded both as a categorical variable (age_group: *2-year-olds, 3-year-olds, 4-year-olds and adults*) and as a continuous variable (age_in_months), so that the dataset could be analysed by means of both Fisher’s Exact Tests/Chi-Square tests and mixed-effects logistic regressions [*productivity ~ age_in_months*verb_familiarity*construction, (1|participants)*]. However, the dataset did not meet the assumptions for regression models. Hence, in order to carry out the analysis of the results, Classification Trees that factored in age as a categorical predictor (age_group) were used instead (see the analysis, chapter 14).

During the test phase, children interacted with the Experimenter (E.), a plush toy mole (*la Talpa Silvia*, “Silvia the Mole”) and the agent of each action (either *Peppa Pig* or *Emily Elefante*). Morpho-syntactic productivity with the training verbs, the control-familiar verb and the test-nonce verb was elicited through questioning via the plush mole.

11.2.2. TRAINING, CONTROL AND TEST VERBS

The following verbs are those to which the subjects were exposed during the experiment. All verb-action pairs are transitive, regular, Conjugation I verbs:

a) Training-familiar verbs:

Lavare “to wash”:

The agent picks up a bucket and acts as though she was spilling water on the patient’s head. The agent then cleans the patient’s head with a little sponge.

Pettinare “to comb”:

The agent approaches the patient and strokes the latter’s head with a small toy hairbrush.

b) **Control-familiar verb:**

Lanciare “to throw”

The patient sits on one end of a catapult, while the agent jumps on the other end, catapulting the former away.

c) Test-nonce verb:

Bodare:

The agent stands 10 cm away from the patient and stares at her. The agent spins and skips around the patient and gets behind her. Once behind the patient, the agent touches the patient’s back with her nose, causing the patient to fall.

Henceforth, the verbs outlined in (a), (b) and (c) shall be referred to as **training verbs**, **familiar verb** and **nonce verb** respectively.

11.2.3. PROPS AND TOYS.

Children interacted with E. and *la Talpa Silvia*, a plush mole who was blind, not too intelligent and continuously fell asleep. E. voiced and moved *Silvia*.

The actors who performed the actions (agent and patient) were two soft toys, representing two characters of the popular cartoon *Peppa Pig*: *Peppa Pig* herself and *Emily Elephant* (in Italian *Emily Elefante*).

A small toy hair brush, a small bucket and a small sponge were also used so that the actions of combing and washing could be performed by Peppa and Emily.

11.2.4. AUDIO-RECORDING AND LOCATION

Both familiarisation and test phase for the children took place at the local nurseries they attended. Each test phase was audio-recorded and the children were tested wherever a space could be provided by the nurseries (see Appendix_III, ch. 21.2). Adult participants were tested in a quiet room at the experimenter's house. One two-year-old subject underwent both familiarisation and test phases at home with his/her mother and siblings in attendance.

11.3. PROCEDURE

11.3.1. FAMILIARISATION: GROUP-PHASE (four hours over two days)

During the first day of familiarisation E. was introduced to the children by the teachers. Afterwards, E. casually played with them and participated in the nurseries' routines for about two hours in order to gain the children's trust.

On the second day, a plush mole who wore sunglasses and whose name was *La Talpa Silvia* "Silvia the Mole" was introduced to the group. The children were told that the animal was blind, a bit silly (*tonta*) and continually fell asleep. The children were told that "*Silvia* normally lives underground and she needs her sunglasses to protect her eyes from the sunlight". Silvia "took part" in the normal daily routine of the nurseries, consisting of chatting, playing, singing etc., during which the children helped her to overcome her blindness-related problems (*where is X*? *What are you doing*? *Can you carry me there? I can't see*). This was meant to familiarise the children with her blindness.

Familiarisation with the mole's narcolepsy and with the task (answering her questions) was obtained in two ways:

a) during the casual play which took place during this phase, Silvia continually fell asleep. When awake, Silvia very frequently asked questions about what had happened whilst she was sleeping (*What happened?; What did you do?*; *What did X do?*).

b) when possible, children were also introduced to the task by means of a twenty-minute puppet show. The mole would get on stage, greet the children and introduce them to *Peppa Pig* and *Emily Elefante*. Afterwards, Silvia would fall asleep and Emily and Peppa would start playing together. Once finished playing, Emily and Peppa would leave the stage and Silvia would wake up and ask *cosa ha fatto la Peppa?* "What did Peppa do?"; *cosa hanno fatto Pepppa ed Emily?* "What did Peppa and Emily do?"; *cosa ha fatto la Emily?* "What did Emily do?" *cosa è successo?* "What happened?". Children were then encouraged to shout the answer.

11.3.2. TEST-PHASE (individual sessions)

It would now be advisable for the reader to examine the schematic representation of the experimental task in fig. 11.1 (p. 301) before continuing to read.

11.3.2.1. Warm-up phase

As soon as E. and the child arrived at the location of the test phase, Silvia fell asleep and E. then said to the child "*well, let's let her sleep. What about if I show you which kinds of games her friends play in the meantime?*"

If the child said yes, E. then added "*Ok, now I'll show you some games, and later on, when Silvia wakes up, you are going to explain to her what has happened. Would you help me to explain to Silvia what happens?*

Afterwards the two toys, *Peppa Pig* and *Emily Elefante*, were introduced to the child and E. made sure that the child both remembered and was able to utter the toys' names. If the child was not able to utter a name or wanted to name a toy differently, E. and the child negotiated a new name for it (see Appendix_III, ch. 21.3.1.2).

11.3.2.2. Stimuli (training and experimental phase)

Each participant was then exposed to the two training verbs, the familiar verb and the nonce verb, in that order. Each verb-action pair was presented to the child in its IMP.2.SG and infinitive form through an **Introduction** and **three or four blocks of stimuli**. Afterwards, production of the same verb in a past (*passato prossimo*) transitive environment was elicited.

The reader may now wish to examine the schematic representation of the experimental task in fig. 11.1 (p. 301) before continuing to read.

In the **Introduction**, the verb was initially introduced to the child with (1) and other further sentences like (3). E. also made sure that the action was understood as transitive (4-5) and that the child was able to articulate the verb in its infinitive form (2). In this phase, the child was exposed to five instances of the verb in its infinitive form and both the child and E. made the agent act upon the patient several times.

1. *Ti facccio vedere una cosa che si chiama verbare.*

   "I'll show you something that is called *verb-are* [verb-INF]."

2. *Prova a dire verbare.*

   "Try to say *verb-are* [verb-INF]."

3. *Ti faccio vedere come si fa a verbare.*

   "I'll show you how to *verb-are* [verb-INF]."

4. *Guarda cosa la [AGENT'S NAME] fa alla [PATIENT'S NAME]!*

   "look at what [AGENT'S NAME] is going to to [PATIENT'S NAME]!"

5. *Hai visto cosa la [AGENT'S NAME] ha fatto alla [PATIENT'S NAME]?*

   "Have you seen what [AGENT'S NAME] did to [PATIENT'S NAME]?"

During **Block 1– Block 3**, the child was exposed to eight imperative forms of the verb, presented within either a "Vo" (7) or a "V,A" (6) construction. The child was also encouraged to order the agent to perform the action by uttering the same stimuli at least five times. Each stimulus was combined with one to three performances of the action. Throughout this phase the child was also exposed to five further infinitive forms of the verb (see Appendix_III, ch. 21.3, for a detailed description of the script followed by E.)

6. **V, A**
   Verba, [AGENT'S NAME]!
   verb-a [AGENT'S NAME]!
   verb-IMP.2.SG [AGENT'S NAME]!
   "Verb, [AGENT'S NAME]!"

7. **Vo**
   Verbala!
   verb-a=l-a!
   verb-IMP.2.SG=clitic.3.ACC-F.SG!
   "Verb her!"

11.3.2.3. Elicitation and second trial

After the child had gone through Introduction-block_3, the mole would wake up and ask *che cosa ha fatto la [AGENT'S NAME]?* "What did [AGENT'S NAME] do?"; this prompted the child to use the verb in a morpho-syntactic environment (*Agent has verbed Patient*) different from the one in which the verb had been encountered (*verb her!* or *verb, Agent!*).

If the child answered using the target verb in the target construction – i.e. either *(S)+AUX<V<O (Agent has verbed Patient)* or *(S)+o<AUX<V* (*Agent her has verbed*), the tasks moved on to the following verb.

If the child did not use the target verb in the target construction, the game moved on differently, depending on whether the answer (or non-answer) pertained to the **training verbs** or the verbs in the **experimental phase (**familiar verb and nonce verb).

If, **during training** (*lavare* "to wash" and *pettinare* "to comb"), the child:

a) used the target verb, but did not provide the target answer, or

b) used a verb different from the target one used in the game, or

c) stayed silent,

s/he was helped to produce the target answer. If needed, E. suggested the answer to the child who was then encouraged to repeat it to Silvia several times. Afterwards, the game moved on to the next verb.

If, during the **experimental phase** (*lanciare* "to throw" and *bodare* "to nonce.verb"), the child:

a) provided a syntactically and morphologically unproductive answer (see coding (11.4) for what is classified as a productive answer), or
b) used a verb different from the target one used in the game, or
c) stayed silent,

a further block of stimuli (*block_4*) was provided.

In **block 4,** two further infinitive and five further imperative forms of the verb were presented to the child who was then encouraged to order the agent to act upon the patient three times using this imperative form. Afterwards, the mole would wake up again and ask *che cosa ha fatto la [AGENT'S NAME]?* "What did [AGENT'S NAME] do?", hence beginning the elicitation process anew. The child's answer was then noted and classified as productive, unproductive or null (see next section). An answer was coded as null whenever the child either used a verb different from the target one and/or remained silent (no answer). Null answers were discarded and not considered for analysis.

11.4. CODING AND RELIABILITY

Each individual test phase was audio-recorded and E. also took notes of the children's answers. A second transcription was made by listening to the recordings and the two transcripts were then compared. In only one case there was a conflict between handwritten notes and the transcribed recordings; this was solved in favour of E.' notes with the help of the unanimous opinions of three volunteers (a primary school teacher, a philosophy graduate and a physicist). Two children's recordings were incomplete and the coding for their answers relied only on the notes taken by E.

The children's answers were coded separately for morphological and syntactic productivity. An answer was considered morphologically productive whenever the children used a verb ending different from the one heard (e.g. different from

IMP.2.SG and infinitive[74]). As for syntactic productivity, an answer was classified as productive whenever the children combined the verb with a DO[75].

Vocabulary size was assessed through two five-minute sessions during which each child took an adapted (shorter) version of Cianchetti and Sannio Facello's (2010) language test (*Test di Valutazione del Linguaggio*: **tvl**, henceforth; see Appendix_III, ch. 21.4).

Figure 11.1 (overleaf) reports a schematic representation of the experimental design.

Appendix_III (chapter 21.3) contains an accurate description of the task the children attended and reports the experimental script followed by E.

---

[74] Since the morpheme *–a* is both the IMP.2.SG and the PRS.3.SG form of Conjugation I verbs, answers which used a verb inflected in its PRS.3.SG were not considered morphologically productive.

[75] See discussion in 11.5.2 for further detail.

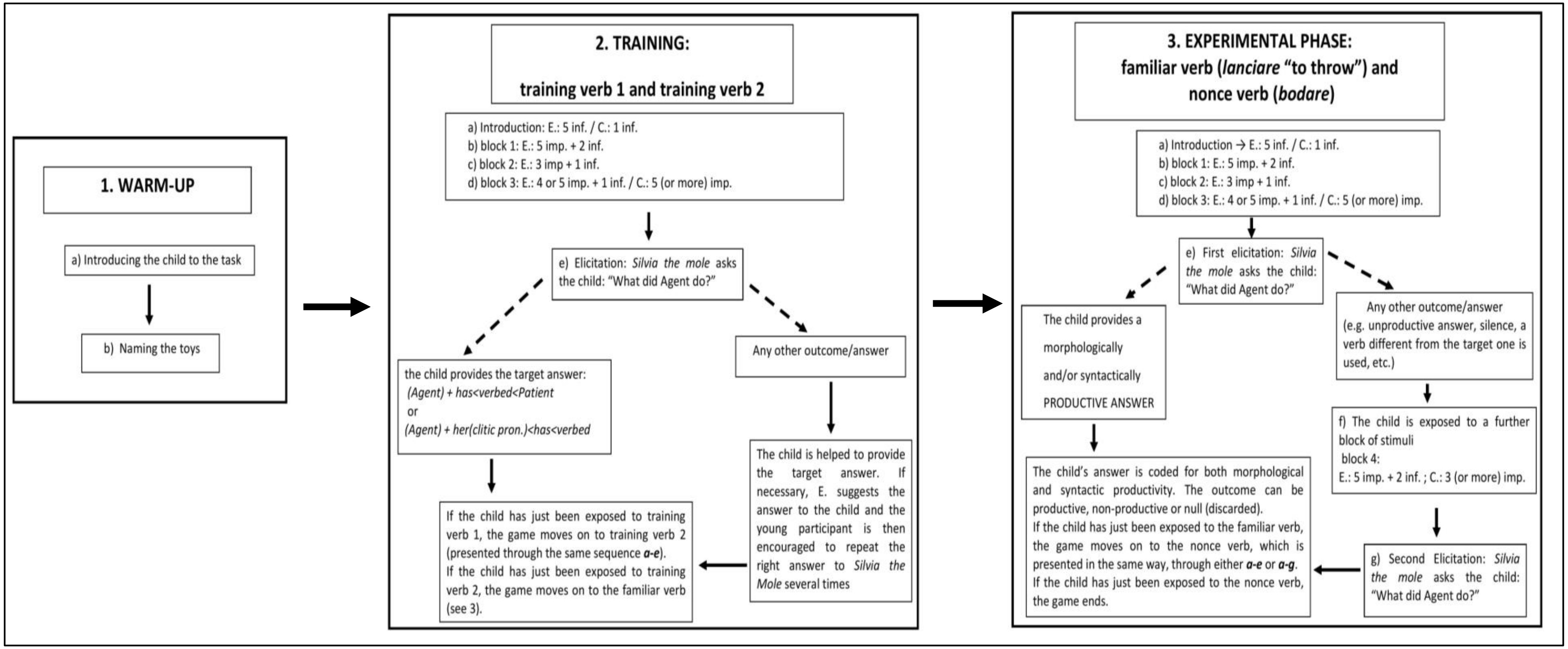


Figure 11.1: the design of the experimental study. Solid arrows indicate the order of the various sections which participants went through. Dashed arrows indicate sequences as consequences of binary possibilities. Hence, after the elicitation trial (2e and 3e), the game would move on differently, depending on the answer (or non-answer) provided by the child. In sections 2 and 3, the letters E. and C. refer to the Experimenter and the child, respectively. "Imp." indicates an imperative stimulus, presented within either a "Vo" or a "VA" construction. "Inf." indicates an infinitive form of the target verb. Hence, in "d) block 3" of both (2) training and (3) experimental phase, the sequence "E.: 4 or 5 imp. + inf./ C.: 5 or more imp." indicates that the Experimenter uttered 4 or 5 imperative forms and one infinitive form of the target verb and that the child uttered 5 or more imperative forms of the same verb. For what is classified as a productive (or non-productive) answer, see the coding section (11.4). For a more detailed illustration of the task, see chapter 21.3 (Appendix_III, Volume II).

11.5. DESIGN RATIONALE

11.5.1. VERB_FAMILIARITY CONDITION

The rationale for exposing participants to both a familiar verb and a nonce verb is probably clear to the reader by now. Productivity with the two types of verbs is (potentially) underpinned by different types of linguistic competence. When prompted to be productive with the familiar verb *lanciare* “to throw”, participants could do so by relying on lexically-specific knowledge of it. When prompted to be productive with the nonce verb, participants would have to go beyond their lexically-specific experience of *bodare*; either by categorising the new expression as an extension vis-à-vis an entrenched lexically-specific unit or by activating a fully-schematic unit.

Exposing participants to the familiar verb provides a control condition to which performance with the nonce verb can be compared: the extent to which participants are equally productive across verb_familiarity conditions can be thought of as the extent to which they can access and retrieve fully-schematic units as easily as lexically-specific ones.

11.5.2. ELICITING MORPHO-SYNTACTIC PRODUCTIVITY: input, output and the construction in which stimuli appeared

As both morphological and syntactic productivity are being investigated, children are prompted to be productive with both. Italian present imperative and declarative *passato prossimo* allow enquiry into both morphological and syntactic productivity by getting children to perform one task only (therefore minimising their efforts).

Children are exposed to verbs presented in their IMP.2.SG form (8a-b) and past tense (*passato prossimo*) transitive uses of the same verbs are then elicited (9a-b).

8. a) “Vo” stimulus

lanciala!
lanci-a=l-a!
throw-IMP.2.SG=clitic.3.ACC-F.SG
“Throw her!”

b) “V,A” stimulus
lancia, Peppa

lanci-a,	*Peppa*
throw-IMP.2.SG,	Peppa

“Throw, Peppa!”

9. a) “(S)+AUX<V<O” output

(Peppa) *ha lanciato (la) Emily*

(Peppa)	*ha*	*lanci-a-to*	*(l-a)*	*Emily*
(Peppa)	has	throw-TV(conj.I)-PTCP	(the-F.SG)	Emily

“(Peppa) has thrown Emily.”

b) “(S)+<o<AUX<V” output
(Peppa) l’ ha lanciata

(Peppa)	*l’*	*ha*	*lanciat-a-t-a*
(Peppa	clitic.3.ACC	has	throw-TV(conj.I)-PTCP-F.SG

“(Peppa) has thrown her.”

In transforming a present imperative sentence into a declarative *passato prossimo* transitive one, children must change the verb’s morphological ending, therefore exhibiting **morphological productivity**: *lancia* (8b) becomes *lanciato* (9a). Furthermore, if the DO is expressed through a clitic pronoun (9b), the past participle ought to agree with the DO (F.SG) in gender and number (*lanci-a-t-a* “throw-TV(conj.I)-PTCP--F.SG)[76]. A sentence like (9b) would be evidence that participants are aware of DO-past_participle agreement when clitic pronouns are used and that they are also able to inflect the past participle of a novel verb for gender and number.

Syntactic productivity is elicited by a design that presents children with stimuli in which verbs appear in two different constructions.

[76] Refer back to 1.1.2.4 and sentence (9).

Under the **“*V,A*” condition**, participants are exposed to transitive_action-verb pairs in which the compulsory object is omitted, but the (optional) agent is expressed in post-verbal position in the form of an addressee (8b).

In order to be conventional (i.e. grammatical and productive) participants must combine the verb (be it novel or familiar) with either an object NP (9a) or a pre-verbal clitic pronoun (9b), therefore providing some information (the patient) that is not attested in the stimulus.

The reader may appreciate that the model (8b) in itself might be misleading, in that children could be brought to infer that the verb is intransitive, as Italian children almost never (0.12%) hear objectless transitive sentences (see Abbot-Smith & Serratrice, 2013). Such an interpretation is however prevented by (4-5) in which E. stresses that the action is transitive (X acts upon Y) and by the prototypically transitive nature of the experimental verbs (there is some kind of force involved in both the familiar verb and the nonce verb).

Under the **“*Vo*” condition**, participants are exposed to transitive_action-verb pairs presented within transitive imperative sentences (Vo; see 8a). Like in the “*V,A*” condition, E. provides cues that the event is one in which X acts upon Y (4-5).

As the reader may recall, clitic pronouns take pre-verbal position when the verb is [+FINITE] (*oV* or *oAUX V*) and post-verbal position (*Vo*) when it is [-FINITE] (compare (8a) and (9b); refer back to 1.3.2).

When prompted to use a *passato prossimo* declarative sentence, participants will have two choices: either moving the stimulus’ clitic before the auxiliary (9b), therefore changing the “Vo” sequence of the stimulus into an “oV” sequence; or substituting it with a full NP (9a). Either way, participants would be showing that they are able to parse the input as *root-IMP.2G=clitic.3.ACC-F.SG*. In the former case (9b), participants would be showing knowledge of clitics’ distributional properties. In the latter (9a), they would be providing evidence that they understand what the clitic maps onto (i.e. the DO).

Naturalistic data suggest that children know clitics’ distributional properties from very early on in their development. Children always correctly place clitics in either pre or post-verbal position, depending on the [+/- FINITE] feature of the verbs with which they combine (Guasti 1993/94; Antelmi, 1997).

If early knowledge of clitics' distributional properties is fully-fledged, children ought to be able to recognise *la* in *lancia**la*** (8a) as a DO. Consequently, they should be able to use the verb in a declarative transitive sentence (9a-b) when the stimulus has "Vo" structure. Unlike under the "V,A" condition, they do not have to provide any missing information.

As for the use of the infinitive form (*-a-re* "-TV(conj.I)-INF"), this is meant to help children to categorise the target verb as a *PROCESS* belonging to Conjugation I.

11.5.3. STIMULI'S ORDER OF PRESENTATION

All children are exposed to two training verbs, one familiar verb and one nonce verb, in that order. This sequence is meant to facilitate the children's task with the nonce verb in two ways.

Firstly, the training phase is meant to help the children understand what the game (the task) is about. The actions combined with the real verbs (training and familiar) are all transitive. This should help the participants to infer that even the nonce verb depicts a scene in which X acts upon Y. This inference is further encouraged by E. verbally emphasising the transitivity of the event (4-5).

Secondly, and most importantly, exposing children to this sequence should facilitate the extension of the syntactic patterns of the real, known verbs (training and familiar) to the unfamiliar, nonce one.

Status of unit is mostly a matter of entrenchment. Initially, children might have a weak representation of the Italian transitive schema. This means that their ability to categorise a novel instance of the transitive construction (*Agent has nonce.verbed Patient*) might initially happen by extension vis-à-vis an entrenched lexically-specific instantiation of it, such as *(DOER)+do<THING*. Children ought to be able to activate their lexically-specific units in order to utter grammatical transitive sentences with known verbs (training and familiar). Units that are fresher in memory (recently activated/retrieved) are supposedly easy to retrieve/activate again (see Dąbrowska, 2004). Hence, exposing children to three

familiar verbs should facilitate the re-activation of those lexically-specific units and therefore facilitate the categorisation of the newly acquired nonce verb as an extension vis-à-vis previously used known, real verbs.

Thus, the sequence *training-familiar-nonce* is meant to facilitate the children's task.

11.5.4. CONTROLLING FOR OBJECT OMISSION

Clitic omission[77] is a well-documented phenomenon in early Italian (Guasti, 1993/1994; Tedeschi, 2007; Serratrice, Sorace & Paoli 2004), which wears off almost completely (3-6% of omission rates) well before the end of the third year of life (2;5–2;7; see Guasti, 1993/1994).

11.5.4.1. The pragmatic nature of object omission

Allen (2000)[78] notes that children's argument omission is influenced by referents' **informativeness** value. When a referent has a low value of informativeness, i.e. it is easily recoverable from both discourse and/or context, children are likely to omit it, even when omission would be ungrammatical (see Serratrice et al., 2004; Tedeschi, 2007).

Tedeschi (2007) experimentally investigated the relationship between

> "clitic omission and the pragmatically uninformative contexts in which clitics are used in adult language".
>
> Tedeschi (2007, p. 204)

She showed children (2;6-6;7) pictures depicting causative events (a man combing a girl's hair) and elicited a description of them. Elicitation could take the form of specific (*What is Agent doing to Patient?*) or generic (*What is happening?)* questions. She predicted that children would drop more objects after specific questions, as both referents had already been activated by the question itself (*What is **Agent** doing to **Patient**?).* The specific question assigns a low value of informativeness to the patient. Conversely, the generic question assigns a high value of informativeness to it.

[77] Clitic omission stays for cases in which children drop the object when the pragmatics of the situation would require the object to be expressed by a clitic.
[78] Mentioned in Tedeschi (2007, pp. 202-203).

The youngest group's (2;6-2;11) answers showed equal object-omission rates (21%), irrespective of question type. The children in the middle group (3;6-4;1) omitted more objects when the question was specific (35%) than when it was general (11%; Wilcoxon Signed Ranks z = -1.826, *p*=0.068; Tedeschi, 2007, p. 210). The older children (5;5-6;5) virtually never produced null objects. The only three cases of null object, though, were attested after specific questions (Tedeschi, 2007; table 5, p. 211).

Hence, children older than 3;6 seem to be sensitive to discourse pragmatics, whereas younger children seem not to be. Tedeschi hypothesises that the youngest children's omissions could be explainable in terms of extra-linguistic factors that make referents active on the scene and hence potential candidates for null realisation. For instance, she points out that

> "it is anyway arguable that the physical presence of the referents in the context can be a sufficient requirement for younger children to assume that the referents are accessible to their hearer".
>
> Tedeschi (2007, p. 212)

The next subsection walks the reader through how the design controls for both linguistic and extra-linguistic factors that may lead to object omission.

11.5.4.2. The mole and her question

The experiment depends on children answering questions asked by *la Talpa Silvia* "Silvia the mole". This design choice has a threefold motivation:

a) Since E. is the one who tells the children what the two toy protagonists are doing, if children were to be questioned by E. as opposed to by a third actor who did not witness what just happend (i.e. the mole), they might be misled to think that something different from the target action (Agent acted upon Patient) is the focus of the question.

b) E. introduces the Patient to the child; this makes the patient's participation to the action part of the shared knowledge between E. and the child. This is likely to assign a low value of informativeness to the patient (hence, it increases the likelihood of DO null realisation).

c) Physical presence of the referent itself might be enough to assign a low value of informativeness to the DO (Tedeschi, 2007).

Making children interact with a blind and narcoleptic mole represents a way around these problems since:

a) the mole is asleep while the two toys are playing and therefore misses what goes on. This creates the pragmatic conditions for the mole enquiring into what has happened, without which the question would be logically odd.

b) the mole sleeps while the toys are performing the action, there is no shared knowledge between her and the child as to who is acted upon and what the nature of the action is. This creates a shared knowledge between the child and the mole in which the DO has a high value of informativeness.

c) the mole's blindness weakens any inferences originating in the physical presence of the patient. If the mole does not see, she cannot be fully aware of the physical presence of the patient. This is meant to increase the informativeness value of the DO.

The question asked by the mole (*che cosa ha fatto la [AGENT'S NAME]?* "what did [AGENT'S NAME] do?") assigns a high value of informativeness to both verb and DO, as it gives them value of NEW (hence, it should lower the likelihood of DO null realisation). It is also specific enough to make sure that children understand that the question is about what just went on between the two toys, as opposed to a generic *cosa è successo?* "What happened?" which might lead children to infer that the question's focus is something different from the target action.

Thus, the design attempted to control for both linguistic (the type of question) and extra-linguistic (the blind and narcoleptic mole) factors behind object omission.

11.6. A FINAL UNBALANCED DESIGN

11.6.1. UNBALANCED NUMBER OF STIMULI

If a child got distracted or wanted to play other games, E. would go along with this whilst slowly and respectfully leading the child back to the task. The task was then re-started at the beginning of the particular block that had been interrupted. So, if the child started to be unresponsive halfway through block_2, once s/he had been brought back to the task, the study would re-start from the beginning of block_2. This procedure was necessary with many two-year-olds and a few older children.

Four children also got tired or bored or unresponsive during the game. In those cases, the test phase was broken down into two sessions.

Consequently, those children who got distracted halfway through the study were exposed to more instances of the target verbs than those children who carried out the task without interruptions.

11.6.2. UNBALANCED CONDITIONS

Not all children contributed to the data for both the nonce verb and the familiar verb. Indeed,

a) six children in the main study avoided using the nonce verb (they used another known verb, instead), therefore contributing to the familiar verb condition only.

b) Because of the high drop-out rate during the main study (about 35%; twenty-seven out of seventy-seven children), it was necessary to conflate data from both the preliminary study and the main study so that a reasonable sample size could be obtained. Due to slightly different choices about which verbs were selected as training and control (familiar verb) in the preliminary study, eight of twelve children who took part in the preliminary study only contributed data for the nonce verb (see Appendix_III, ch. 21.5 for details) .

Furthermore, whenever a child either could not or did not want to utter either toy's name, E. negotiated a different name for the toy(s) with him/her.

These facts had three consequences:

a) Some children contributed to data for both the familiar verb and the nonce verb, whereas others contributed data for one or the other only (table 11.2).

b) The construction in which verbs appeared ("Vo" and "V,A") ended up being not equally distributed across each *verb_familiarity X age_group* combination (table 11.3).

c) There is not an equal distribution of agent types (*Peppa Pig* or *Emily Elefante*) across each combination of age_groupXverb_familiarityXconstruction (table 11.4).

Table 11.2: participants and the experimental conditions to which they contributed.

age group	gender		no of participants in each age group	participants' contribution			no of answers for the familiar verb	no of answers for the nonce verb
	m	f		familiar and nonce verb	familiar verb only	nonce verb only		
two-year-olds	9	7	16	9	4	3	13	12
three-year-olds	11	5	16	15	1	0	16	15
four-year-olds	9	12	21	15	1	5	16	20
adults	5	9	14	14	0	0	14	14

Table 11.3: the distribution of construction conditions across verb_familiarity and age_group.

age_group	familiar verb		nonce verb	
	"Vo"	"V,A"	"Vo"	"V,A"
two-year-olds	7	6	4	8
three-year-olds	6	10	10	5
four-year-olds	9	7	9	11
adults	7	7	7	7

Table 11.4: the distribution of agent and construction conditions across verb_familiarity and age_group.

verb_familiarity	familiar verb						nonce verb					
agent	Peppa_Pig		Emily_Elefante		chosen by the participant		Peppa_Pig		Emily_Elefante		chosen by the participant	
construction	"Vo"	"V,A"	"Vo"	"V,A"	"Vo"	"V,A"	"Vo"	"V,A"	"Vo"	"V,A"	"Vo"	"V,A"
two-year-olds	3	4	0	1	4	1	2	5	0	1	2	2
three-year-olds	2	6	4	3	0	1	6	2	3	2	1	1
four-year-olds	5	3	4	3	0	1	4	4	4	7	1	0
adults	3	4	4	3	0	0	3	4	4	3	0	0

12.

PREDICTIONS

To my knowledge there are no experimental studies which have investigated Italian-speaking children's linguistic production with nonce verbs (but see Abbot-Smith & Serratrice (2103) on comprehension). The following predictions are therefore based on Italian children's behaviour with real linguistic material but in light of previous studies that have employed nonce verbs to investigate the acquisition of English.

12.1. MORPHOLOGICAL PRODUCTIVITY

Akhtar and Tomasello's (1997, study 4) found that English-speaking children become productive with the past tense suffix *–ed* only after they turn three years of age. Previous studies (Caselli et al, 1999; Devescovi et al., 2005) showed that Italian children are morphologically more advanced than their English-speaking peers. This is the result of children's sensitivity to their ambient language, as morphology plays a much bigger role in Italian than in English. Hence, if English-speaking three-year-olds are morphologically productive with nonce verbs, so should Italian ones be. Three and four-year-olds' morphological productivity should therefore be adultlike across verb_familiarity conditions.

As for two-year-olds, previous studies (Pizzuto & Caselli, 1994; Leonard, Caselli & Devescovi, 2002) suggest that verbs' paradigms are not fully mastered by children younger than three years of age. Two-year-olds are therefore expected to be less productive than older participants. However, since productivity with the familiar verb could also be obtained out of lexically-specific knowledge, their performance as a group might still be adultlike (i.e. not statistically different from that of adults). Conversely, productivity with the nonce verb is linked to a more schematic representation that two-year-olds might not have developed yet. Hence, their performance with *bodare* is expected to be statistically poorer than that of older participants, particularly adults.

Overall, two-year-olds should perform statistically better with the familiar verb than with the nonce verb. All other age-groups should perform statistically similarly across verb_familiarity conditions.

12.2. SYNTACTIC PRODUCTIVITY

All age groups should perform similarly with the familiar verb, for which even the youngest children could rely on lexically-specific knowledge of *lanciare* “to throw”.

Were the results consistent with previous studies on English-speaking children (see Tomasello, 2000b, 2006b; Akhtar, 1999), the number of participants productive with the nonce verb should increase as a function of age, with four-year-olds performing better than three-year-olds, who in turn should perform better than two-year-olds. Were the results consistent with the outcomes of Akhtar’s (1999) study, four-year-olds, but not the younger age-groups, should perform in an adultlike manner.

Furthermore, the two younger age-groups are also expected to perform statistically better with the familiar verb than with the nonce verb. They should also be more likely to be productive with the nonce verb when the stimulus explicitly presents linguistic cues of transitivity (“Vo” condition) than when it does not (“V,A” condition).

12.3. VOCABULARY, AGE AND PRODUCTIVITY

Previous studies (Marchman & Bates, 1994; Caselli et al., 1999; Bates & Goodman, 2001; Devescovi et al. 2005) found that vocabulary is a better predictor of grammatical development than age. These findings are consistent with usage-based models (refer back to 3.9.2). Goldberg (1999) hypothesises that as learners acquire more and more words (and the patterns in which they appear), re-organising the system into more schematic units might be a way to lighten storage memory. Since productivity with the nonce verb is hypothesised to be a by-product of the extent to which participants can access more schematic units, vocabulary is expected to be a better predictor of productivity than age, irrespective of verb_familiarity (familiar *versus* nonce) and type of productivity (morphological *versus* syntactic).

13.

RESULTS

Table 13.1 and figure 13.1 summarise the results. The following sections illustrate the results pertaining to morphological (13.1) and syntactic (13.2) productivity separately.

Table 13.1: Results by age group. As discussed in chapter 11, some children contributed data for only one or the other verb_familiarity condition, hence the unequal number of answers between nonce verb and familiar verb (refer back to table 11.2).

morphology						
age group	familiar verb			nonce verb		
	prod	not prod	tot	prod	not prod	tot
two-year-olds	8	5	13	4	8	12
three-year-olds	15	1	16	11	4	15
four-year-olds	16	0	16	16	4	20
adults	13	1	14	14	0	14
syntax						
age group	familiar verb			nonce verb		
	prod	not prod	tot	prod	not prod	tot
two-year-olds	8	5	13	2	10	12
three-year-olds	14	2	16	6	9	15
four-year-olds	16	0	16	14	6	20
adults	13	1	14	13	1	14

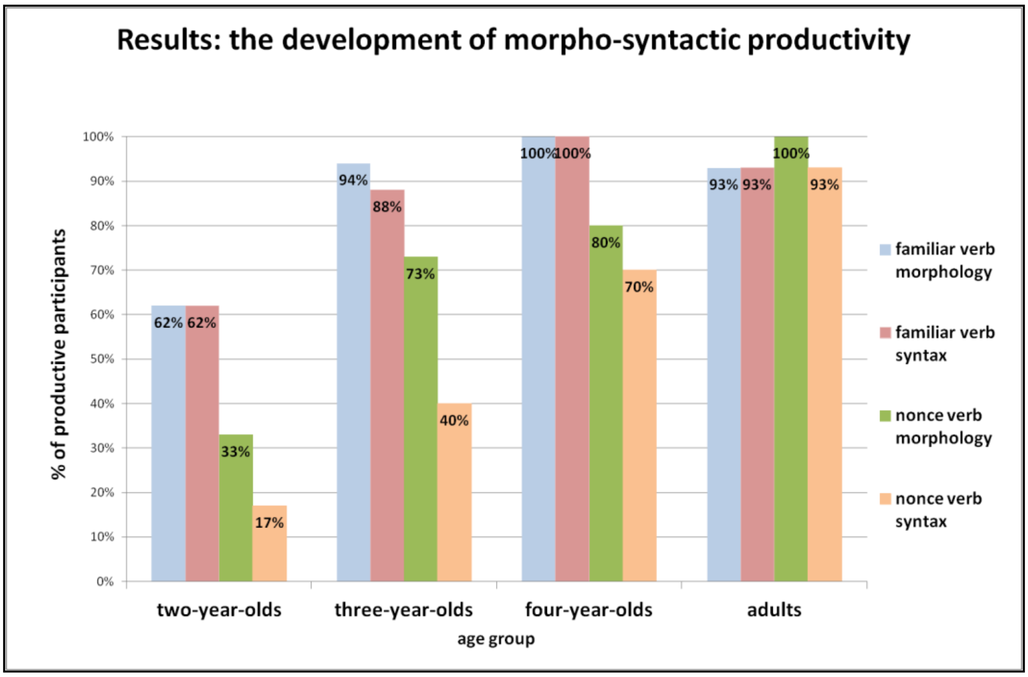


Figure 13.1: morphological and syntactic productivity: summary of results.

13.1. MORPHOLOGICAL PRODUCTIVITY

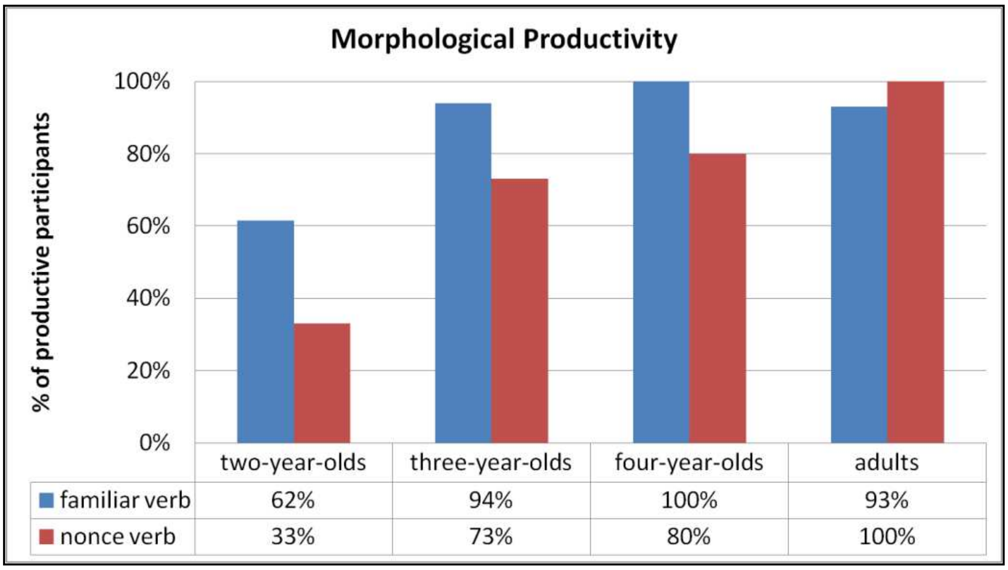


Figure 13.2: morphological productivity as a function of age_group and verb_familiarity.

Fig. 13.2 shows that all children's groups performed better with the familiar verb than with the nonce verb. **Two-year-olds** performed worse than the other age-groups with both the familiar verb (62% of productive participants vs. 93-100%) and nonce verb (33% of productive participants vs. 73-100%). The gap between two-year-olds and three-year-olds is slightly bigger in the nonce (-40%) than in the familiar (-32%) verb_familiarity condition. Overall though, most of the youngest children were productive with the familiar verb (62%), but unproductive with the nonce one (33%).

Virtually all participants in the remaining age-groups were morphologically productive with the **familiar verb** (93-100% of productive participants), and both children's groups performed slightly better (94%-100% of productive participants) than adults (93%[79]). Of the fifty-two participants who were morphologically productive, fifty uttered at least one sentence in which the verb was conjugated in its past participle form. Two three-year-olds showed a morphologically productive behaviour by inflecting the verb in its PRS.1.SG form (*lanci-o* "throw-PRS.1.SG").

[79] This is because one adult used a causative construction (*La Peppa ha fatto lanciare la Emily*; "Peppa has made Emily throw") instead of the target transitive construction (*La Peppa ha lanciato Emily*; "Peppa has thrown Emily"). Since the verb was neither combined with a DO nor inflected, this was not considered a productive answer, either morphologically or syntactically.

The proportion of participants who were productive with the **nonce verb** increases as a function of age: 33% (two-year-olds), 73% (three-year-olds), 80% (four-year-olds) and 100% (adults). All productive participants uttered at least one past participle form (*bodato*) of the nonce verb (*bodare*).

13.2. SYNTACTIC PRODUCTIVITY

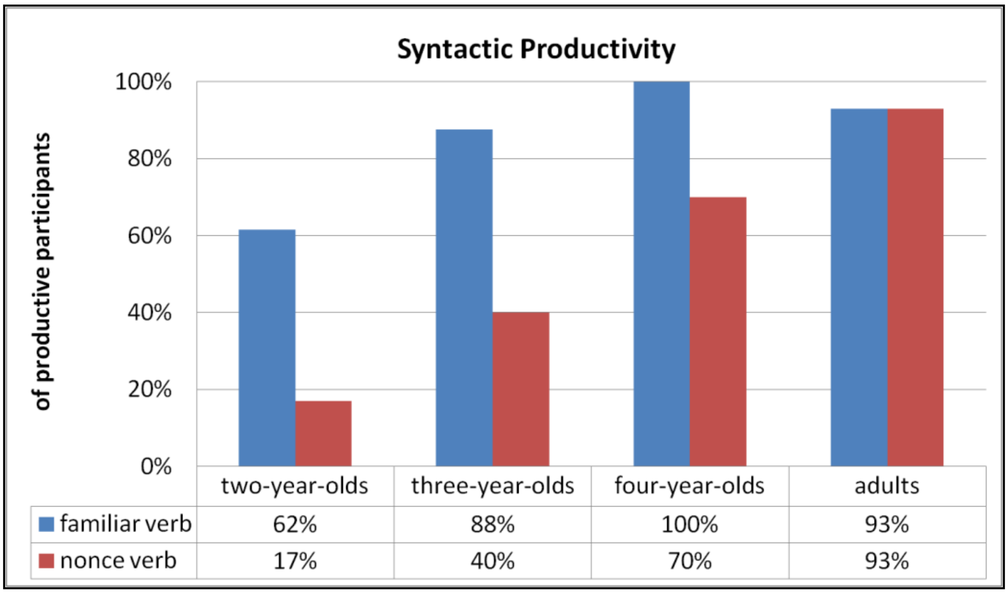


Figure 13.3: syntactic productivity as a function of age_group and verb_familiarity.

Even syntactic productivity (fig. 13.3) shows that children performed better with the familiar verb than with the nonce verb. **Two-year-olds** were less productive than the other age-groups, with both the familiar verb (62% vs. 88-100% of productive participants) and the nonce verb (17% vs. 40-93%). Within this group, the gap between familiar verb and nonce verb (-45%) is bigger than with morphology (-32%): syntactically, two-year-olds were nearly three times as productive with the familiar verb (62%) as with the nonce verb (17%).

As for the other age-groups, participants were mostly productive with the **familiar verb**, although three-year-olds performed slightly worse (88% of productive participants) than four-year-olds (100%) and adults (93%).

Performance with the **nonce verb** shows a gradual and steady increase in productivity as participants get older. Three-year-olds were more than twice as productive as two-year-olds (17% of productive participants), but still performed poorly (40%). Four-year-olds performed better and were mostly productive (70% of productive participants). Finally, adults' performance was nearly at ceiling (93% of productive participants[80]).

Fig. 13.4 shows the results pertaining to syntactic productivity by factoring in the construction with which stimuli were presented.

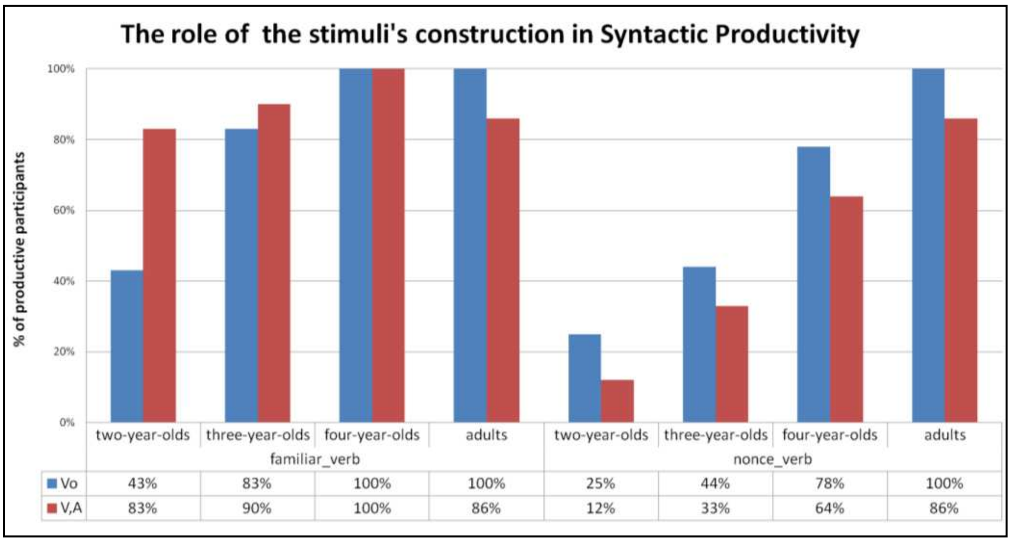


Figure 13.4: syntactic productivity as a function of age_group, verb_familiarity and construction.

In each age-group there were more syntactically productive participants when the **nonce verb** *bodare* was presented in a transitive frame (Vo) (25%-100% of productive participants) than when it had "V,A" structure (12%-86%).

Results pertaining to the familiar verb show that two-year-olds, and to a minor extent three-year-olds, were more productive when stimuli had "V,A" structure. Indeed, the former were overwhelmingly productive with "V,A" stimuli (83% of productive participants), but not with "Vo" stimuli (43% of productive participants). Four-year olds performed at ceiling in both conditions, but one adult was not productive when the stimulus presented "V,A" structure (hence 86% of productive participants amongst adults).

[80] The only adult who was not productive uttered an intransitive construction (*ha bodato* "(she) has nonce.verbed") instead of the expected transitive.

14.

ANALYSIS

14.1. THE EXPERIMENTAL DESIGN:

productivity ~ age*verb_familiarity*construction

For both morphology and syntax, two mixed-effects (Baayen, 2008; Bates, 2010; Johnson, 2008) backwards (Larson-Hall, 2010a; Crawley, 2007) logistic regressions were fitted to the dataset using the package lme4 of the software R (https://cran.r-project.org/web/packages/lme4/index.html). For both morphology and syntax, a model was built with productivity as a dichotomous DV (productive vs. not_productive) and participants as the only random effect. Predictors were age_in_months (continuous), verb_familiarity (dichotomous, *familiar* vs. *nonce*) and construction (dichotomous, *Vo* vs. *V,A*). The effect size of the best model was computed by calculating each model's R2 using the MuMIN package (https://cran.r-project.org/web/packages/MuMIn/index.html) (see also Nakagawa & Schielzeth, 2013). However, an inspection of the correlation tables showed that both datasets (morphology and syntax) presented multicollinearity, with some correlations as high as 0.97, which obviously yielded unrealistic odds ratios. Factoring in participants' age-group as a ranked variable (2 to 5) did not solve the problem in a satisfactory way. Furthermore, the dataset on morphology presented problems of quasi-perfect separation.

Analysis was therefore carried out using Classification Trees (Hothorn, Hornik & Zeileis, 2006; Strobl, Mallesy & Tutz, 2009; Berk, 2006). Classification and Regression trees (CARTs, henceforth) are powerful, non-parametric methods that:

a) allow for predictors to be highly correlated,
b) make no assumptions as to the distribution of the dataset,
c) can handle datasets in which there are many predictors (be they categorical or continuous) and a small number of subjects,
d) can detect both linear and non-linear relationships.

The reader not acquainted with the method can refer to chapter 21.6 of Appendix_III (part V, Volume II), where a brief and simplified illustration of CART is provided by presenting a summary of Strobl et al.'s (2009) and Berk's (2006) works.

The choice of CARTs as method of analysis also presents the following advantages:

a) Their results are very straightforwardly interpretable (Strobl et al., 2009; Boyd & Goldberg, 2012).

b) They make it possible to factor in variables that are highly correlated, which solves the aforementioned multicollinearity problems (Strobl et al., 2009; Berk, 2006).

c) CARTs do not make assumptions as to the distribution and nature of the learning sample. This means that it is not necessary to factor in age as a ranked predictor (2 to 5) and that age can be factored in as a categorical predictor (age_group). This is how it has been factored in in previous studies on English-speaking children (Akhtar & Tomasello, 1997; Akhtar, 1999; Abbot-Smith et al., 2001). Thus, factoring in age as age_group (*2-year-olds*, *3-year*-olds, *4-year-olds* and *adults*) allows a more straightforward cross-linguistic comparison of results. Finally, it means that the results pertaining to morphological productivity can be analysed without dropping those variables that bring about quasi-perfect separation.

d) CARTs can analyse datasets in which there are many predictors and a small number of subjects. This makes it feasible to include in the model even those variables that were not controlled for. For instance, it is possible to investigate whether there is a difference in productivity between children who participated in the main study and children who were tested in the preliminary study, whether children's gender had any impact on results and whether children who attended a specific nursery were more productive than children who attended another one. That many effects (and their interactions) would be very problematic to handle for regression models based on 120 observations (see Strobl et al., 2009).

e) When it comes to spotting interactions, CARTs often outperform regression models (see Yo, Ference, Cote & Schwartz, 2012).

For both morphological and syntactic productivity, a Classification Tree was grown adopting the unbiased classification algorithm[81] (Hothorn et al., 2006). Hothorn et al.'s (2006) method was chosen because it allows *R* to set a tree-growth stopping criterion based on the strength (*p*-value) of the association between predictors and DV (i.e. it avoids pruning). Each tree had either syntactic or morphological productivity as dichotomous DV (*productive* vs. *not_productive*). The predictors included were:

age_group: participants' age group (2_year_olds, 3_year-olds, 4_year_olds and adults)

verb_familiarity: whether answers were provided with either the familiar verb (familiar_verb) or the nonce verb (nonce_verb).

construction: whether stimuli were presented with "Vo" (vo) or "V,A" (va) structure.

pre_vs_main: whether participants' answers were collected during the preliminary (pre) or main (main) study.

school: the nursery attended by the children (coded as A, B, C, D and E) and, for adults, their level of education.

gender: whether participants were male (m) or female (f).

The tree-growth stopping criterion was set at $p \geq 0.30$ and p-values were adjusted with Bonferroni corrections.

Fig. 14.1 and 14.2 report the results for morphological and syntactic productivity, respectively.

[81] available in the R package *partykit* (https://cran.r-project.org/web/packages/partykit/index.html)

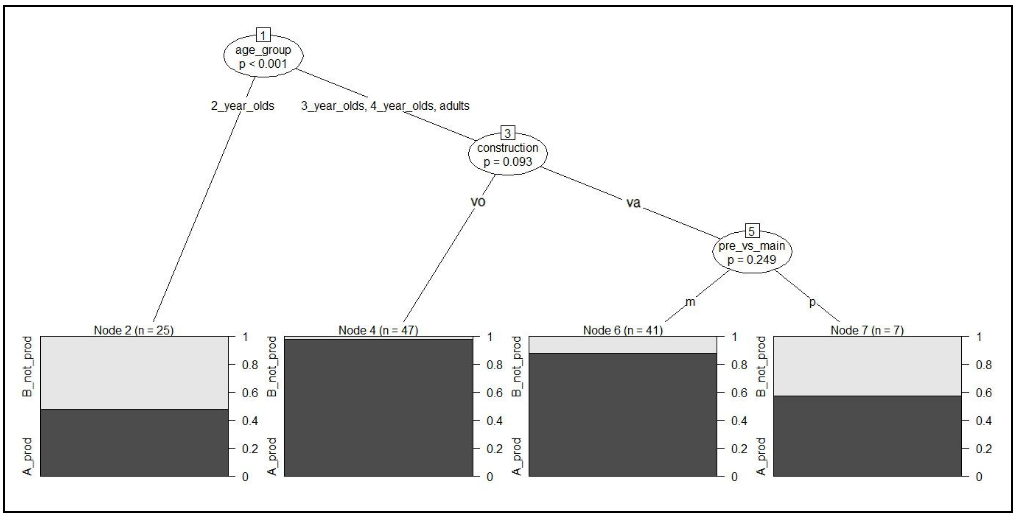


Figure 14.1: Results pertaining to morphological productivity.

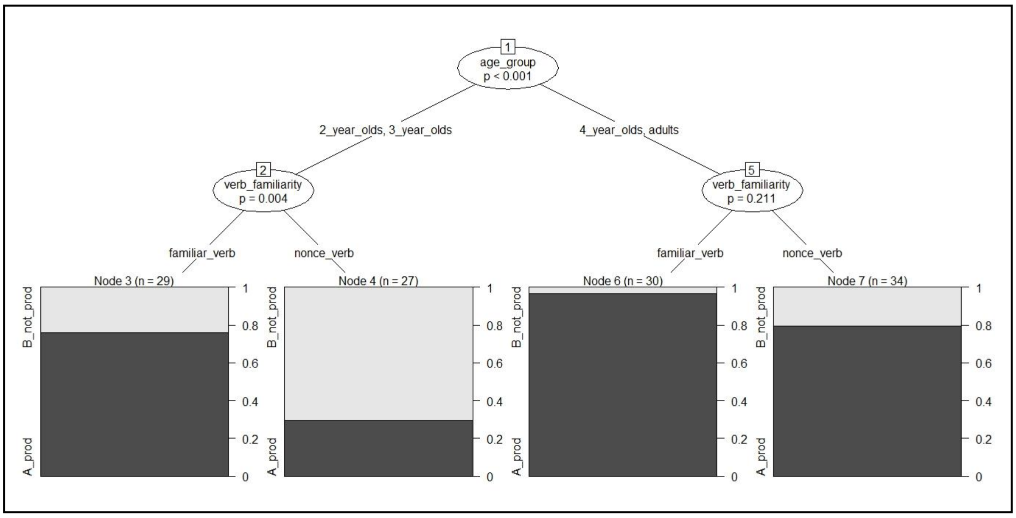


Figure 14.2: Results pertaining to syntactic productivity.

As for **morphological productivity,** the only predictor that reaches significance (p<.001) is age_group, in that two-year-olds were less productive than the other age groups considered together, irrespective of the verb_familiarity condition (fig. 14.1, node 1). However, participants' performance was not affected by verb_familiarity, as no effect of it is seen in fig. 14.1. Hence, when considered as a group, even two-year-olds appear to be capable of inflecting known and newly learnt verbs with not too dissimilar mastery (p≥.3).

As for **syntactic productivity**, age_group is still the strongest predictor in that there were fewer productive two-year-olds and three-year-olds than four-year-olds and adults (fig. 14.2, node 1, p<.001). The two younger groups performed much worse with the nonce verb than with the familiar verb (30% vs. 76% of productive participants; fig. 14.2, node 2, p=.004). Conversely, the proportion of adults and four-year-olds who were productive with the nonce verb is similar to the proportion of them who were productive with the familiar verb (fig. 14.2, node 5, p=.21). Construction is not selected by the tree.

14.1.1. GENERALISING BEYOND THE POPULATION SAMPLED

Strobl et al. (2009) warn that the predictive power of CARTs can show high variability as their structure represents a very tight description of the particular learning sample on which they are based. For example, the selection of the splitting variable (e.g. age_group rather than verb_familiarity) and the exact position of a given cut-off point (two-year-olds vs. all other age-groups) are highly dependent on the specific sample analysed. As a result, the outcome structure could be substantially different if the learning data were even slightly different. Such a problem is known as overfitting, i.e., the final structure not only captures those structural features that could be generalisable to the whole population, but also those structural features that are due to random variation (Strobl et. al., 2009).

Hence, in order to investigate whether results would be generalisable to the whole population, outcomes were further analysed by means of Fisher's Exact Tests and by means of eight mixed-effects logistic regressions, two for each age_group (syntax and morphology). Models had productivity as dichotomous DV, verb_familiarity as predictor and participants as the only random effect[82]. Since construction is a significant predictor of neither morphological (fig. 14.1) nor syntactic (14.2) productivity, the analyses focussed on the interaction between verb_familiarity and age_group.

Table 14.1 presents the results of regression models. Table 14.2 shows all the pairwise comparisons between all age-groups with respect to the familiar verb and nonce verb using Fisher's Exact Tests (none of the matrixes met the assumptions for analysis with Chi-Square Tests).

[82] By considering each age-group separately and by not factoring in *construction*, it is possible to "cheat" a way out of multicollinearity problems. Nevertheless, problems of quasi-perfect separation remain. It follows that results of these models have to be taken with caution and only as a means of validating other tests' results. Since estimates were calculated with nAGQ=10 (syntax) and nAGQ=15 (morphology), it was not possible to use the package MuMIN to calculate models' R2.

The next sections discuss the results of all statistical tests in a unified manner and set their outcomes against the predictions put forward in chapter 12.

Table 14.1: Mixed-effects logistic regressions. For each model, productivity is the dichotomous (productive vs. not_productive) DV and verb_familiarity the dichotomous predictor (familiar vs nonce). The only random effect is participants.

MIXED-EFFECT MODELS: productiviy ~ verb_familiarity + (1|participants)				
(a) MORPHOLOGICAL PRODUCTIVITY				
age_group	Intercept Estimate (Std. Error)	Intercept z-value	verb_familiarity (familiar vs nonce) Estimate (Std. Error)	verb_familiarity (familiar vs nonce) z-value
two-year-olds	2.07 (2.94)	0.70	-4.68 (4.54)	-1.03
three-year-olds	**2.71 (1.03)**	**2.62****	-1.32 (1.22)	-1.09
four-year-olds	22.95 (648.86)	0.03	-21.51 (639.30)	-0.03
adults	2.565e+00 (1.038e+00)	2.47	3.459e+01 (1.794e+07)	0
(b) SYNTACTIC PRODUCTIVITY				
age_group	Intercept Estimate (Std. Error)	Intercept z-value	verb_familiarity (familiar vs nonce) Estimate (Std. Error)	verb_familiarity (familiar vs nonce) z-value
two-year-olds	0.47 (0.57)	0.82	***-2.08 (0.96)***	***-2.16*****
three-year-olds	**15.23 (6.97)**	**2.16****	***-18.17 (8.22)***	***-2.21*****
four-year-olds	30.31 (2053)	0.015	-26.44 (2063)	-0.014
adults	**2.565e+00 (1.038e+00)**	**2.47****	-2.387e-15 (-2.387e-15)	0

*p<.1; **p<.05

Table 14.2: Pair-wise comparisons across age groups with the familiar verb (left) and the nonce verb (right) with respect to morphological (upper tables) and syntactic (bottom tables) productivity. *P*-values are adjusted with Bonferroni corrections (*p<.1 and **p<.05).

MORPHOLGY - FAMILIAR VERB: Fisher's Tests	adults	four-year-olds	three-year-olds	two-year-olds
adults				
four-year-olds	0.467 (0 - 34.13)			
three-year-olds	0.87 (0.01 - 73.16)	Inf (0.03 - Inf)		
two-year-olds	7.52 (0.68 - 412.02)	Inf (1.38 - Inf)*	8.66 (0.78 - 470 .74)	

MORPHOLGY - NONCE VERB: Fisher's Tests	adults	four-year-olds	three-year-olds	two-year-olds
adults				
four-year-olds	Inf (0.48 - Inf)			
three-year-olds	Inf (0.67 - Inf)	1.09 (0.22 - 9.59)		
two-year-olds	**inf (3.62 - Inf) ****	7.39 (1.25 - 55.41)	5.12 (0.82- 39.75)	

SYNTAX - FAMILIAR VERB: Fisher's Tests	adults	four-year-olds	three-year-olds	two-year-olds
adults				
four-year-olds	0 (0 - 34.13)			
three-year-olds	1.82 (0.09 - 117.77)	Inf (0.19 - Inf)		
two-year-olds	7.52 (0.68 - 412.02)	Inf (1.38 - Inf)*	4.14 (0.53 - 53.21)	

SYNTAX - NONCE VERB: Fisher's Tests	adults	four-year-olds	three-year-olds	two-year-olds
adults				
four-year-olds	5.33 (0.53 - 275.15)			
three-year-olds	**17.4 (1.76 - 912.82) ****	3.37 (0.70 - 18.03)		
two-year-olds	**48.73 (4.03 - 294.60)****	10.66 (1.59 - 129.40)*	3.19 (0.42 - 40.28)	

* p < .1   ** p < .05

14.1.2. THE DEVELOPMENT OF MORPHOLOGICAL PRODUCTIVITY

The most relevant finding is that verb_familiarity is not a significant predictor (p≥.30). Indeed, such a DV does not "make it to the tree" (fig. 14.1) and all regression models (table 14.1a) suggest that, for each age-group, the proportion of productive participants is similar across verb_familiary conditions (p>.05). This suggests that, contrary to predictions, as a group, Italian-speaking two-year-olds (M=2;08, range=2;2-3;0) appear to have acquired the schema *PROCESSa-to* "PROCESS-TV(conj.I)-PTCP"[83].

Predictions also hypothesised that two-year-olds would be less productive than the other participants. The tree-analysis (node 1, fig. 14.1) is consistent with these predictions in that the proportion of productive two-year-olds is significantly (p<.001) smaller than the proportion of productive participants in the other age-groups considered *together* (two-year-olds **vs.** three-year-olds *plus* four-year-olds *plus* adults). However, when the youngest group's performance is compared with each older children's age group (Fisher's Exact Tests; table

[83] Or, better put, the proportion (33%) of two-year-olds who use the past participle with the nonce verb (and hence that can be said to be morphologically productive) is not statistically (p≥.30) smaller than the proportion of two-year-olds (62%) who are productive with the familiar verb (with which lexically-specific knowledge would suffice to carry out the task successfully). Such an outcome is further discussed in terms of graded entrenchment of schematic constructions in 14.1.4 and 16.1.4.1.

14.2), their (group) performance is statistically similar to that of any other children's age-group, irrespective of the verb_familiarity condition.

Finally, the results of the Fisher's Exact Tests (table 14.2) show that the proportion of productive two-year-olds (62%) was statistically similar to the proportion of productive adults (93%) with the familiar verb (p>.05, Bonferroni correction), but not with the nonce verb (33% vs. 100%; fig. 13.2; p<.05, Bonferroni correction), which is in line with predictions. However, the extremely high odds ratio and upper CI (both Inf.) cast doubt as to whether this could be a case of Type I error. Be that as it may, the results appear to be accountable for in terms of piecemeal entrenchment of the schematic unit *PROCESSato*. Two-year-olds were (statistically) as productive with the familiar verb as with the nonce verb, which could be interpreted as evidence of ability to retrieve the schema *PROCESSato* "PROCESSed" to categorise new expressions. This means that by the age of 2;08, the schema is entrenched enough to be exploited (evoked) in production tasks with new material, at least in a proportion of two-year-olds that is not statistically smaller than the proportion of them who cannot be said to be morphologically productive. However, the two-year-olds' schema is not as entrenched as that of adults as it is inferred on the basis of a much poorer linguistic experience; hence adults' better performance. As children get older (three and four years old), the schema becomes entrenched enough to yield adultlike productivity[84].

14.1.3. THE DEVELOPMENT OF SYNTACTIC PRODUCTIVITY

Before discussing the developmental path that emerges, I shall provide a unified overview of the results yielded by the statistical tests adopted: classification tree

[84] One could argue that the non-statistical significance between productive participants with familiar verb (62%) and nonce verb (33%) here is an indication that participants at age two are just as "unproductive" with *both* types of verbs. That is, they had equal problems with both the familiar verb and nonce verb. However, that they were unproductive with the familiar verb is not true, since most (62%) of two-year-olds were morphologically productive with *lanciare* 'to throw' and indeed, as a group, they were as statistically productive as adults. Hence, as a group, they *were* productive with the familiar verb. It is however true that only very few of them (33%) were morphologically productive with the *nonce* verb (hence, as a group, two-year-olds were not productive with the nonce verb). However, the noteworthy outcome is that they were *not statistically* less productive with the nonce verb than with the familiar verb. This outcome appears to suggest that Italian-speaking two-year-olds can retrieve (morphological, word-level) schematic and lexically-specific constructions with not dissimilar mastery (p>.05). See Conclusion (chapter 15) for further discussion.

(fig. 14.2), mixed-effects logistic regressions (table 14.1b) and Fisher's Exact Tests (table 14.2).

14.1.3.1. A unified overview of all statistical tests

Fisher's Exact Tests (table 14.2) show that all age groups performed similarly (p>.05) with the familiar verb. As for the nonce verb, both two-year-olds and three-year-olds were significantly (p<.05, Bonferroni correction) less productive (17% and 40% of productive participants, respectively) than adults (93% of productive participants). Conversely, four-year-olds performed in an adultlike manner (p>.05). These outcomes perfectly mirror predictions and are consistent with previous findings regarding English-speaking children (e.g. Akthar, 1999).

Nevertheless, node 1 in fig. 14.2 shows that the proportion of productive two-and-three-year-olds was (significantly, p<.001) smaller than the proportion of productive four-year-olds and adults (two-year-olds *plus* three-year-olds **vs.** four-year-olds *plus* adults), irrespective of the verb_familiarity condition. Such a discrepancy can be explained by considering the bi-partitioning nature of CARTs. The method works in terms of recursive dichotomous splits, each one yielding a node. The predictor that is selected first is the one that more strongly relates to the DV. Once a variable (age_group, in this case) is selected, the method dichotomously splits the dataset so as to group together similar values of the response variable (productive vs. not_productive) on the basis of the values attested in the chosen predictor (age_group). In other words, it establishes the node that most reduces impurity (see Strobl et al., 2009). Hence, what the tree shows is that two-year-olds and three-year-olds *together* performed worse (54% of productive answers) than four-year-olds and adults *together* (88% of productive answers), when the results pertaining to both familiar verb and nonce verb are considered *together*. Fisher's Exact Tests and node 1 are therefore compatible and consistent with each other, as they show different comparisons. The former (table 14.2) compares age-groups with each other on familiar verb and nonce verb, the latter (fig. 14.2) compares the two youngest age-groups against the two oldest age-groups on productivity with both types of verbs. Fisher's Exact Tests (table 14.2) show that such a difference in productivity is due to performance with the nonce verb, rather than to performance with the familiar verb. With the latter, 76% of two-year-olds and three-year-olds were syntactically productive, as opposed to 97% of four-year-olds and adults. With the nonce verb, only 30% of participants in the two youngest groups were syntactically productive, whereas 79% of four-year-olds and adults used *bodare* productively. Indeed, node 2 (fig. 14.2) shows that the two youngest age-groups were significantly (p=.004) more productive with the familiar verb than with the nonce verb. This is consistent with predictions and is borne out by the regression models (table 14.1b) which show that both two (z=-2.16, p=.031) and three (z=-2.21, p=.027) year-olds were, as a group, significantly more productive with the familiar verb than with the nonce verb. Conversely, both four-year-olds and adults were similarly productive across verb_familiarity conditions (fig. 14.2; table 14.1b)

14.1.3.2. A piecemeal development

The developmental path that emerges shows a gradual increase in the proportion of syntactically productivity participants with both the familiar verb (62%, 88%, 100%, 93%) and the nonce verb (17%, 40%, 70%, 93%). Furthermore, the gap in productivity between the two verbs reduces as a function of age (-45%, -44%, - 30%) and disappears during adulthood (0%) (fig. 13.3). This fits the general picture depicted by Tomasello (2000b, 2006b) who reports that performance with nonce verbs shows a gradual and steady improvement up to the ninth year of life (refer back to fig. 3.9).

Ultimately, results can be interpreted as being a function of the gradual entrenchment of the constructional schema *(S)+V<O*. As the schema becomes more and more entrenched (i.e. its status of units strengthens), it also becomes easier to retrieve in order to categorise novel expressions. The result is an increasing proportion of participants who are productive with newly learnt material (the nonce verb) as a function of age. Schematic representation appears to be a matter of degree, rather than and in/out distinction: it develops slowly and gradually throughout the early years.

Such a gradual development can be better grasped by looking at tables 14.1b and 14.2. Performance on the familiar verb is similar across age-groups (table 14.2); this suggests that when linguistic experience provides lexically-specific, concrete examples, even the youngest children can behave linguistically in an adultlike manner.

However, when two-year-olds and three-year-olds have to use newly learnt verbs in syntactic environments that are new for those verbs, they perform significantly more poorly than adults (table 14.2) and also significantly worse than they do with known material (table 14.1b). This can be interpreted as evidence that the Italian transitive construction is not entrenched enough before the fifth year of life. Put simply, two-year-olds and three-year-olds' syntactic competence is better describable as lexically-specific, rather than fully-fledged.

The very gradual development (none of the pair-wise comparisons between children's age-groups reaches significance) leads to a small, yet crucial improvement between the fourth and the fifth year of life (+30%, p=.58). Four, but not three, year-olds perform in an adultlike manner (table 14.2) and also manage to perform equally well with both familiar verb and nonce verb (table 14.1b). This suggests that four, but not two and three, year-olds have acquired the Italian declarative transitive constructional schema, of which they appear to have an adultlike representation.

As a final remark, it has to be noted that, as predicted, the proportion of syntactically productive two-year-olds and three-year-olds was higher when the nonce verb was encountered within a transitive frame (Vo) than when the stimulus was objectless (V,A) – fig. 14.3. However, *construction* is not selected by the tree in fig. 14.2, which means that such an effect is highly insignificant ($p \geq .30$). Because of that, such an outcome is not further discussed.

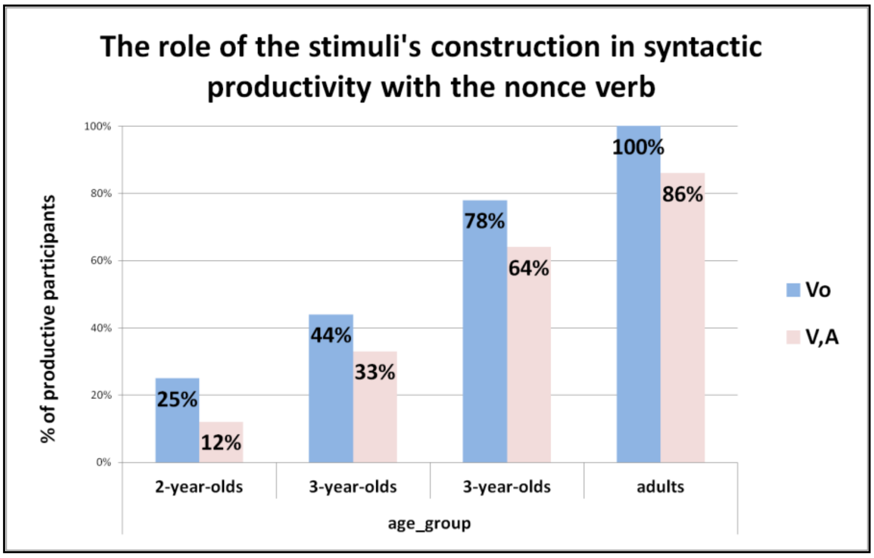


Figure 14.3: proportion of syntactically productive participants with the nonce verb depending on the construction with which stimuli were presented.

14.1.4. ON THE EARLY EMERGENCE OF MORPHOLOGICAL PRODUCTIVITY

Taken as a group, Italian-speaking two-year-olds (2;02-3;00) appear to start showing productivity with regular Conjugation I past participles (according to the statistical analysis, at least; table 14.1; fig. 14.1) by the age of 2;08 (2;02-3;00). If results are compared with Akhtar and Tomasello's (1997) findings, Italian-speaking two-year-olds appear to be more productive with past participles than their English-speaking peers[85]. Thus, it may be the case that the Italian schema *PROCESSato* is acquired earlier than its English counterpart *PROCESSed*. Such an outcome is consistent with previous studies' findings (Caselli et al., 1999; Devescovi et al., 2005) which showed that, when paired for age and vocabulary, Italian-speaking children are morphologically more advanced than English-speaking children.

Such a different pace of acquisition can be explained in terms of language-specific factors. Form-function mapping is likely to be easier for the Italian Conjugation I past participle *PROCESSato* than for its English counterpart *PROCESSed*.

[85] Only three out of twenty (15%) English-speaking children aged 2;01-2;09 in Akhtar and Tomasello's (1997) study were productive with the *-ed* morpheme, as opposed to four out of twelve (33%) of this study's children who were productive with the schema *PROCESSato*.

Languages grammaticalise functions with different means/forms. Children acquire those forms by attending to the input to which they are exposed and mapping them onto their functions. Patterns of morphological inflection (be they verbal or nominal) are word-level schemas (Booij, 2010; refer back to 2.2.7).

Token frequency facilitates entrenchment and therefore representational strength and status of unit (Dąbrowska, 2004; Langacker, 2000). The more entrenched a pattern (be it schematic or lexically-specific), the easier it is to activate. Type frequency is thought to help generalisation processes and hence the development of more schematic units (Tomasello, 2003). If a pattern, be it partially (*where's THING?*) or fully (*AGENT-PROCESS-PATIENT*) schematic, shows a semantically coherent and wide set of instantiations (*where's Mark?*, *where's the ball?*, *where's mum?*), functional analogy will be facilitated, as the same pattern is instantiated by different items that have the same function (Lieven & Tomasello, 2008). As a result, grammatical functions that are coded by very frequent constructions (high token frequency) which can also be instantiated by a wide variety of items (high type frequency) should be (relatively) easy to acquire. That is, ease of schema abstraction (form-meaning mapping) can be thought of as a function of the interaction between token and type frequency.

The mapping of a form onto a function/meaning may be easier in some languages than in others (see Bates & MacWhinney, 1987). For example, Turkish-speaking children learn to disambiguate agent/patient role quite early on in development (around the age of two) because they can rely on clear, frequent and regular case marking at the word level. Turkish cases are incredibly regular (no morpho-phonetic subclasses), phonologically salient (they are stressed), have transparent form-function mappings (no gender contrasts), are very frequent (high token frequency) and apply to virtually all nouns (high type frequency) (Tomasello, 2003; Devescovi & D'Amico, 2001). Hence, the Turkish case system represents an ideal set of (coherent) strings from which children can generalise lexically-bound schemas, such as [[THING]-[ACCUSATIVE]/[...]<[i]], by the age of 2;6 (see Slobin & Bever, 1982). When languages map the same functions onto more complex and less regular patterns (lower type frequency), the acquisition process is not as fast. For instance, the nightmarish case system of Russian is not fully mastered before the ninth or tenth year of life (Devescovi & D'Amico, 2001). Such differences in age of acquisition can be regarded as by-products of form-function mapping transparency in the ambient language.

In Italian, about 70% of verbs belong to Conjugation I with only four of these Conjugation I verbs (and their derivatives) being irregular (refer back to 1.1.2.1). This means that:

a) the schema *PROCESSato* scores high in both token and type frequency, as it is the past participle form of most Italian verbs.

b) when a verb belongs to Conjugation I, there are virtually no less regular, competing forms the verb could take.

Furthermore, the past participle inflection *–ato* is bi-syllabic (/a/ + /to/) and stressed (/bo'd**a.to**/). Hence, the inflection is phonologically salient and easily perceivable (see Bates & MacWhinney (1987) on the role of phonological saliency in facilitating form-function mapping).

On the contrary, English verbs are not grouped into conjugational classes and past participles can take many irregular forms. Furthermore, the past participle morphemes (/t/, /d/, /ɪd/) are unstressed and hence they are not particularly phonologically salient. These factors may be behind the later acquisition of the schema *PROCESSed*. Indeed, when the morpheme to be acquired is highly regular (no other competing forms), frequent (high token frequency), syllabic (hence phonologically salient) and applies to a wide range of items (high type frequency), such as the case of the progressive inflection *–ing*, even English-speaking children are productive from a very young age (about 2;1 - see Akhtar & Tomasello, 1997).

Although there are many irregular past participles in Italian as well, only four of these belong to Conjugation I. Both the IMP.2.SG *–a* and the infinitive *-are* markers are typical of Conjugation I. The experimental stimuli then provided clear cues that the nonce verb (*bodare*) belonged to the largest and most regular inflectional class. Consequently, its participle could take no form other than *bodato*.

Importantly, the fact that (33% of) two-year-olds were able to categorise the stimulus as a *PROCESS* belonging to Conjugation I also implies that those children acquired either (or most likely, both) the IMP.2.SG schema *PROCESSa* or (and) the infinitive schema *PROCESSare*. This in turn may suggest that those children:

a) were able to parse the experimental stimuli as *PROCESS-IMP.2.SG*, *PROCESS-INF*, *PROCESS-IMP.2.SG=HER*
b) could link the schema *PROCESSato* with either (or both) *PROCESSare* "to PROCESS" and/or *PROCESSa*/*PROCESSa=la* "PROCESS-IMP.2.SG/PROCESS-IMP.2.SG=her".

Importantly, being able to link various word-level schemas to each other implies that, at least to some extent, a more schematic construction that subsumes them (*PROCESS-INFLECTION*) is becoming entrenched. Indeed, linking those constructions to each other means that some similarities between them are perceived: this amounts to schema extraction (in Langacker's (2008) terms) [86].

Overall, results suggest that even the youngest children (or at least some of them) have developed generalisations that link different verb-based schemas and thus that they have started mastering a network of constructions at the morphological level.

[86] This hypothesis is further discussed in terms of gradual entrenchment of schematic constructions in section 16.1.4.1.

14.2. AGE, VOCABULARY AND LINGUISTIC PRODUCTIVITY

Devescovi et al. (2005) found that, although both vocabulary and age are positively related to grammatical development, the former is a better predictor than the latter. This section investigates the extent to which vocabulary and age can be said to predict linguistic productivity in early Italian and whether either one is a better predictor than the other.

The children's vocabulary was assessed by testing them through an adapted (shorter) version of Cianchetti and Sannio Facello's (2010) language development test (see Appendix_III, ch. 21.4). Data on vocabulary was not fully available for some participants who were therefore excluded from the analysis. These included:

a) One child who did not attend any vocabulary session

b) Five children who attended only the comprehension session

c) All adult participants whose vocabulary was not tested

This left forty-seven children, aged 2;02 to 5;00.

Previous studies (Marchman & Bates, 1994; Bates & Goodman, 2001; Caselli et al., 1999; Devescovi et al., 2005) investigated the relationship between age in months, vocabulary and grammatical development. Thus, in order to allow comparisons with previous findings, children's age was calculated in months. From now on, I shall refer to children's age as **age_in_months** and to their vocabulary scores as **vocabulary_tvl**.

14.2.1. THE RELATIONSHIP BETWEEN VOCABULARY AND AGE

Figure 14.4 shows that as children get older, their vocabulary scores increase. Since an inspection of the data through histograms showed that neither age_in_months (Shapiro Wilk W = .95, p=.06) nor vocabulary_tvl (Shapiro Wilk W=.94, p=.017) was normally distributed, strength and direction of their relationship were investigated by means of a Kendall's tau correlation test (Field, Miles & Field, 2012). The outcome indicated that vocabulary_tvl and age_in_months are moderately and positively related (τ =.53, p(two-tailed)<.00001).

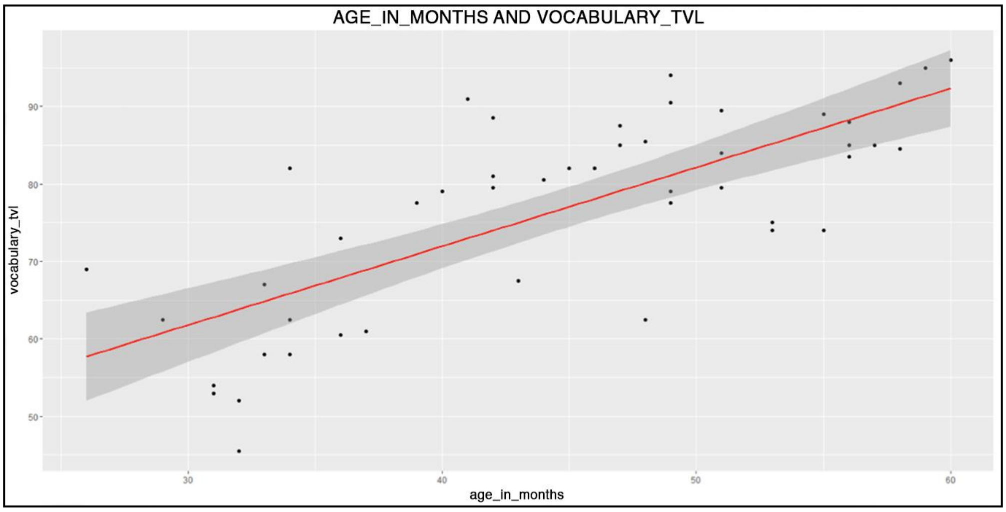


Figure 14.4 The relationships between age_in_months (x-axis) and vocabulary_tvl (y-axis).

14.2.2. VOCABULARY AND AGE AS PREDICTORS OF MORPHO-SYNTACTIC PRODUCTIVITY

This sub-section explores whether age_in_months or vocabulary_tvl is a better predictor of syntactic and morphological productivity by means of point bi-serial correlation tests. Since correlation tests assume independence of data, analyses are presented for familiar verb and nonce verb separately, as verb_familiarity is a within-participants condition. Following Field et al.'s (2012) guidelines, point bi-serial tests were performed by assigning a value of 0 to non-productive answers and 1 to productive ones and by running Pearson's tests between productivity and the two continuous variables under analysis (vocabulary_tvl and age_in_months).

Whether productive children were older and/or had higher vocabulary scores than non-productive children was analysed by carrying out robust independent t-tests, using a 20% trimmed means and percentile bootstrapping (nboot=2000) method, available in the WRS R-package (https://cran.r-project.org/web/packages/WRS2/index.html) (see Larsson-Hall, 2010a).

14.2.2.1. Morphological Productivity

Fig. 14.5 and 14.6 graph morphological productivity as a function of age_in_months and vocabulary_tvl, respectively. Table 14.3 reports the results of all robust t-tests pertaining to morphological productivity. Table 14.4 reports the results of both normal and partial correlations pertaining to morphological productivity.

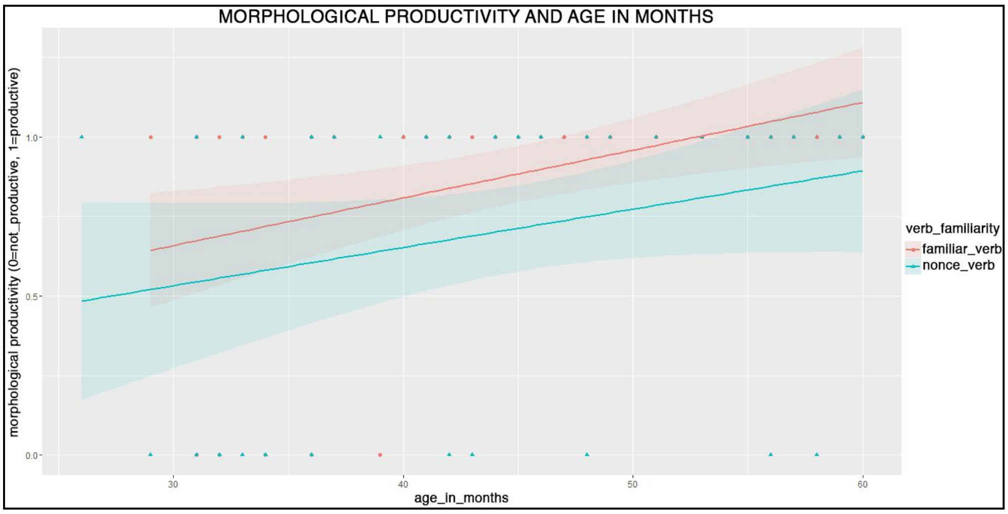


Figure 14.5 : morphological productivity with familiar and nonce verbs as a function of age_in_months.

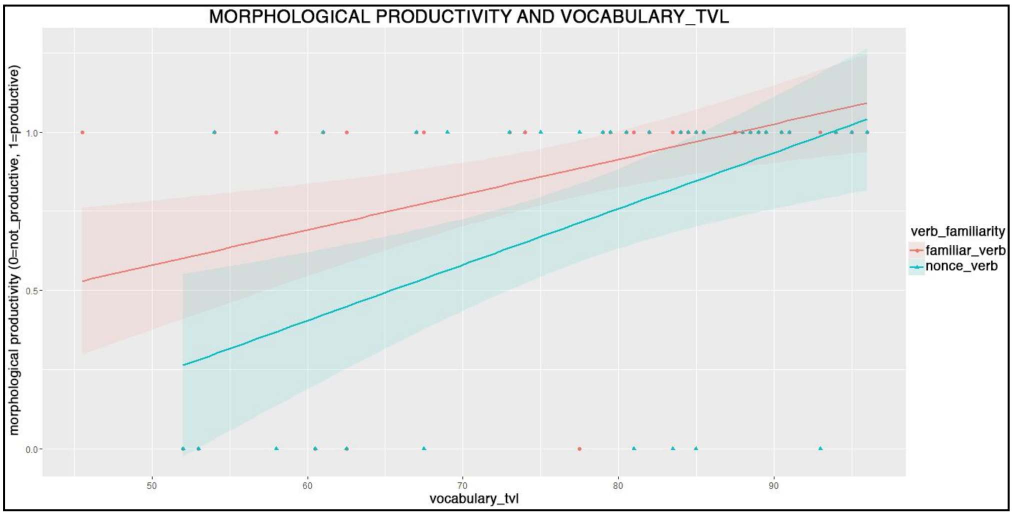


Figure 14.6: morphological productivity with familiar and nonce verbs as a function of vocabulary_tvl.

Table 14.3: Robust t-tests that compare vocabulary_tvl and age_in_months between morphologically productive and morphologically non-productive children. For each subtable (age in months and vocabulary score), Column 1 (lines 2 and 3) indicates verb familiarity. Column 2 reports the number (N=) of children who were productive, the 20% trimmed mean of their age in months (top subtable) and their vocabulary scores (bottom subtable). Column 3 reports the number (N=) of children who were not productive, the 20% trimmed mean of their age in months (top subtable) and their vocabulary scores (bottom subtable). Column 4 reports the difference in trimmed means between productive and non-productive children (and their CIs) regarding the production with familiar verb (line 2) and nonce verb (line 3).

MORPHOLOGICAL PRODUCTIVITY: comparing productive and non-produtive children (t-tests)			
age in months			
verb_familiarity	productive participants: N (20%) trimmed mean	non-produtive participants: N (20%) trimmed mean	difference between (20%) trimmed means (CIs)
familiar verb	***N=38***	***N=5***	***13.13****
	47	***43***	***(7.88 - 17.13)***
nonce verb	N=30	N=12	6.72
	47.22	40.05	(-2.25 - +14.13)
vocabulary score (tvl)			
verb_familiarity	productive participants: N (20%) trimmed mean	non-produtive participants: N (20%) trimmed mean	difference between (20%) trimmed means (CIs)
familiar verb	***N=38***	***N=5***	***22.94****
	81.60	***58.67***	***(8.83 - 29.73)***
nonce verb	***N=30***	***N=12***	***15.03****
	88.28	***67.25***	***(5.25 - 23.47)***

* $p<.05$

Table 14.4: point bi-serial correlations and partial point bi-serial correlations between morphological productivity and either age_in_months or vocabulary_tvl. Results significant at *p*< .05 are *in bold and Italics*. In each sub-table (a and b), the relationship between productivity and either age_in_months (left) or vocabulary_tvl (right) is reported. In each sub-table, it is reported whether results pertain to productivity with the familiar verb (line 2) or the nonce verb (line 3). Columns 2 and 6 report the test statistics (r_{pb}), columns 3 and 7 report the R-squared of the relationship, columns 4 and 8 report the % of variance explained by either age_in_months or vocabulary_tvl and columns 5 and 9 report the *p*-value of the test statistics.

MORPHOLOGICAL PRODUCTIVITY (correlation analyses)								
(a) correlations								
verb_familiarity	age_in_months				vocabulary_tvl			
	r_{pb}	R-squared	explained variance	p-value	r_{pb}	R-squared	explained variance	p-value
familiar verb	***.425***	***.180***	***18%***	***.005***	***.452***	***.204***	***20%***	***.002***
nonce verb	.247	.061	6%	.115	***.474***	***.225***	***23%***	***.002***
(b) partial correlations								
verb_familiarity	age_in_months (vocabulary_tvl held constant)				vocabulary_tvl (age_in_months held constant)			
	r_{pb}	R-squared	explained variance	p-value	r_{pb}	R-squared	explained variance	p-value
familiar verb	.143	.020	2%	=.367	.221	.049	5%	.160
nonce verb	-.221	.045	5%	=.185	***.460***	***.211***	***21%***	***.002***

Children who are morphologically productive with the **familiar verb** are older *and* they have higher vocabulary scores than non-productive children. Children who are morphologically productive with the **nonce verb** have higher vocabulary scores, but are *not* significantly older than children who are not productive (table 14.3).

As for morphological productivity with the **nonce verb, vocabulary_tvl** is clearly a better predictor than age_in_months. The latter is significant after neither normal nor partial correlation analysis. Vocabulary_tvl is moderately related to morphological productivity after both normal (r_{pb} = .474, Rsquared=.225, p=.002) and partial (r_{pb} = .460, Rsquared=.211, p=.002) correlations, explaining 21% of unique variance (table 14.4b).

As for morphological productivity with the **familiar verb** (table 14.4a), vocabulary_tvl and age_in_months seem to be indistinguishable, although vocabulary_tvl accounts for slightly more variance (20%) than age_in_months (18%). When partial correlations that hold the other predictor constant are carried out, neither IV is significant at p<.05. However, Vocabulary_tvl, accounts for more unique variance (5%) than age_in_months (2%).

Growing a classification tree that has morphological productivity with the familiar verb as dichotomous DV and age_in_months and vocabulary_tvl as predictors (fig. 14.7) suggests that the latter is a better predictor than the former. Age_in_months does not reach significance (fig. 14.7, node 2, p=.38), whereas vocabulary_tvl is selected by the method as the first and most significant predictor (fig. 14.7, node 1, p=.007).

Overall, it is possible to conclude that vocabulary is a better predictor of morphological productivity than age with both the familiar verb and the nonce verb.

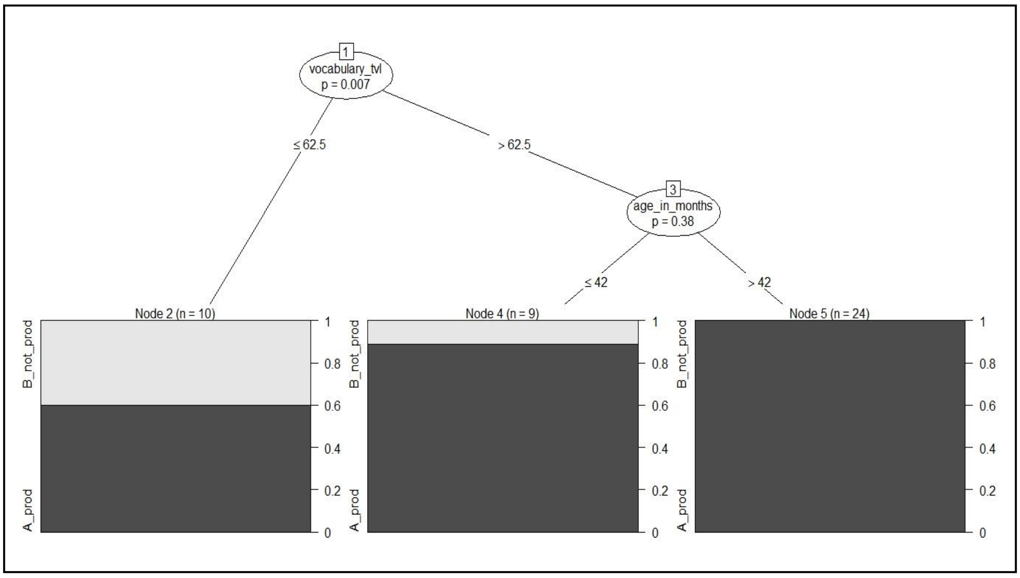


Figure 14.7: morphological productivity with the familiar verb as a function of vocabulary_tvl and age_in_months.

14.2.2.2. Syntactic Productivity

Fig. 14.8 and 14.9 graph syntactic productivity as a function of age_in_months and vocabulary_tvl, respectively. Table 14.5 reports the results of all robust t-tests pertaining to syntactic productivity. Table 14.6 reports the results of both normal and partial correlations pertaining to syntactic productivity.

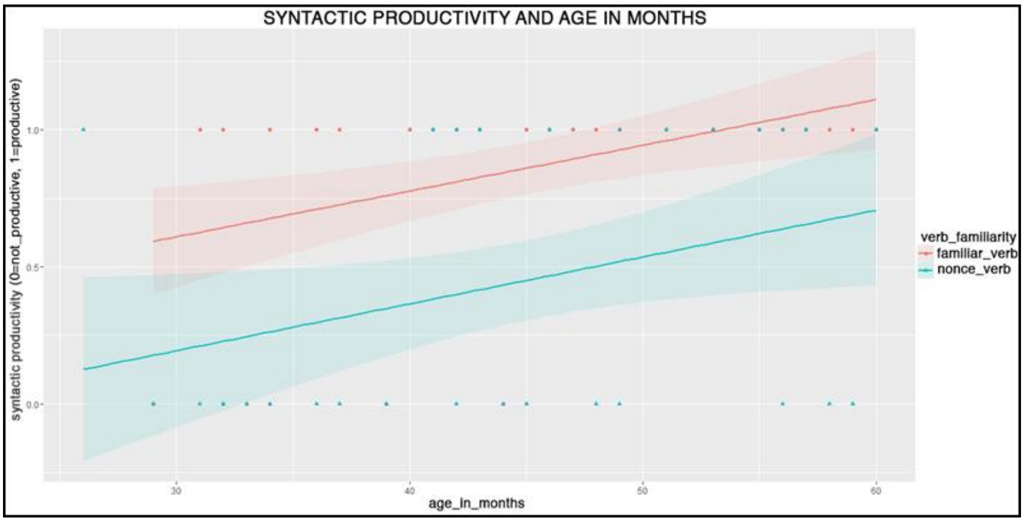


Figure 14.8: syntactic productivity with familiar and nonce verbs as a function of age_in_months.

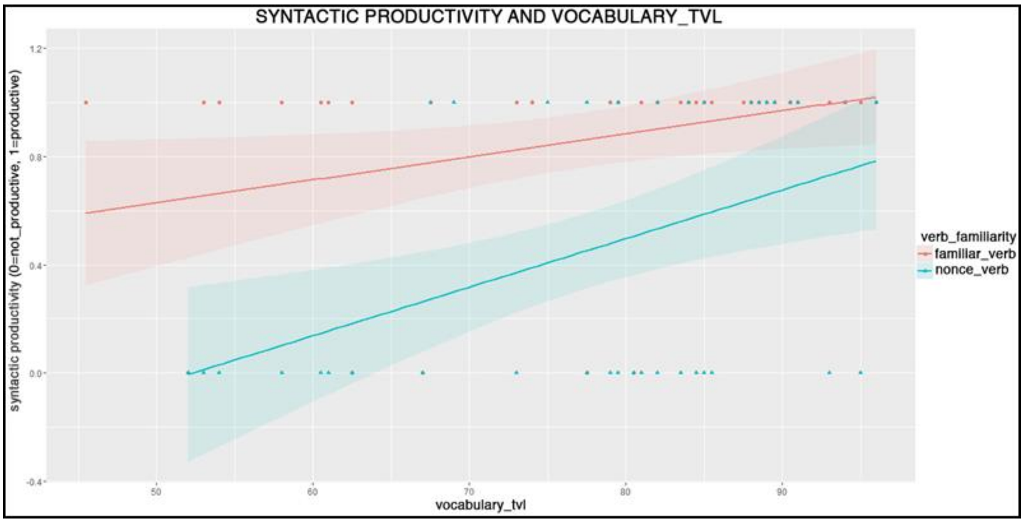


Figure 14.9: syntactic productivity with familiar and nonce verbs as a function of vocabulary_tvl.

Table 14.5: Robust t-tests that compare vocabulary_tvl and age_in_months between syntactically productive and syntactically non-productive children. For each subtable (age in months and vocabulary score), Column 1 (lines 2 and 3) indicates verb familiarity. Column 2 reports the number (N=) of children who were productive, the 20% trimmed mean of their age in months (top subtable) and their vocabulary scores (bottom subtable). Column 3 reports the number (N=) of children who were not productive, the 20% trimmed mean of their age in months (top subtable) and their vocabulary scores (bottom subtable). Column 4 reports the difference in trimmed means between productive and non-productive children (and their CIs) regarding production with familiar verb (line 2) and nonce verb (line 3).

SYNTACTIC PRODUCTIVITY: comparing productive and non-productive children (t-tests)			
age in months			
verb_familiarity	productive participants: N (20%) trimmed mean	non-produtive participants: N (20%) trimmed mean	difference between (20%) trimmed means (CIs)
familiar verb	***N=37***	***N=6***	***12.76****
	47.26	***34.5***	***(6.18 - 17.48)***
nonce verb	***N=19***	***N=23***	***7.69****
	49.16	***41.47***	***(0.90 - 12.90)***
vocabulary score (tvl)			
verb_familiarity	productive participants: N (20%) trimmed mean	non-produtive participants: N (20%) trimmed mean	difference between (20%) trimmed means (CIs)
familiar verb	***N=37***	***N=6***	***14.28****
	81.65	***67.38***	***(3.68 - 23.93)***
nonce verb	**N=19**	**N=23**	**10.94***
	84.04	**73.10**	**(3.11- 18.41)**

* p<.05

Table 14.6: point bi-serial correlations and partial point bi-serial correlations between syntactic productivity and either age_in_months or vocabulary_tvl. Results significant at *p*<.05 are *in bold and Italics*. In each sub-table (a and b), the relationship between productivity and either age_in_months (left) or vocabulary_tvl (right) is reported. In each sub-table, it is reported whether results pertain to productivity with the familiar verb (line 2) or the nonce verb (line 3). Columns 2 and 6 report the test statistics (r_{pb}), columns 3 and 7 report the R-squared of the relationship, columns 4 and 8 report the % of variance explained by either age_in_months or vocabulary_tvl and columns 5 and 9 report the *p*-value of the test statistics.

SYNTACTIC PRODUCTIVITY (correlation analyses)								
(a) correlations								
verb_familiarity	age_in_months				vocabulary_tvl			
	r_{pb}	R-squared	explained variance	p-value	r_{pb}	R-squared	explained variance	p-value
familiar verb	***.437***	***.191***	***19%***	***.003***	***.319***	***.102***	***10%***	***.037***
nonce verb	***.319***	***.101***	***10%***	***.040***	***.437***	***.191***	***19%***	***.004***
(b) partial correlations								
verb_familiarity	age in months (vocabulary_tvl held constant)				vocabulary_tvl (age_in_months held constant)			
	r_{pb}	R-squared	explained variance	p-value	r_{pb}	R-squared	explained variance	p-value
familiar verb	***.317***	***0.100***	***10%***	***.041***	-.020	-.000	0%	.899
nonce verb	-.032	-.001	0%	.842	***.318***	***.101***	***10%***	***.043***

Children who are syntactically productive are older *and* they have higher vocabulary scores than children who are not productive, irrespective of the verb_familiarity condition (table 14.5).

The outcomes of the partial correlation analyses (table 14.6) show that **vocabulary_tvl** (r_{pb} = .318, Rsquared=.101, p=.043) is a better predictor of syntactic productivity with the **nonce verb** than age_in_months (r_{pb} = - .032, Rsquared=.001, p=.842). When the effect of age_in_months is held constant, vocabulary_tvl accounts for 10% of unique variance (table 14.6b). Conversely, **age_in_months** (r_{pb} = .317, Rsquared=.100, p=.041) is a better predictor of syntactic productivity with the **familiar verb** than vocabulary_tvl (r_{pb} = .-0.20, Rsquared=.00, p=.899). Age_in_months accounts for 10% of unique variance (table 14.6b). The classification tree in fig. 14.10 neatly shows how age_in_months and vocabulary_tvl influence children's syntactic productivity.

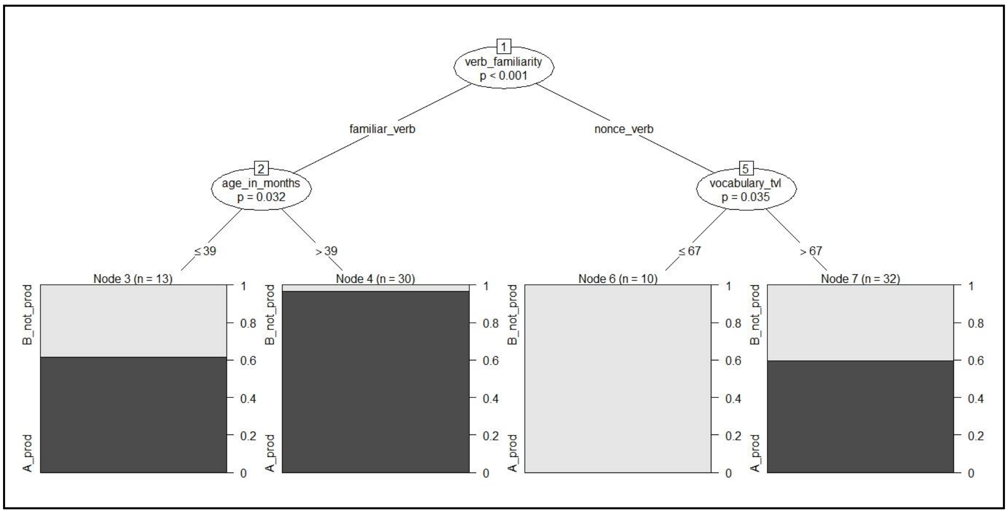


Figure 14.10 syntactic productivity as a function of vocabulary_score, age_in_months, verb_familiarity, construction, pre_vs_main, school and gender, when only children whose vocabulary data are available are considered.

14.2.3. DISCUSSION AND CONCLUSION

The following subsections discuss the results pertaining to the roles of age and vocabulary in the development of morpho-syntactic productivity and set the results against the predictions made in chapter 12.

14.2.3.1. Morphological Productivity, Vocabulary and Age

The results pertaining to morphological productivity are in line with predictions. Vocabulary is a better predictor of morphological productivity than age with both the nonce verb (table 14.4b) and the familiar verb (fig. 14.7), which is consistent with Bates and colleagues' previous findings (Marchman & Bates, 1994; Bates & Goodman, 2001; Caselli et al., 1999; Devescovi et al., 2005). Importantly, this study enquired into an age range (2;2 to 5;0) which is much wider than the age range previously investigated by Bates and colleagues (0:10 to 2;8).

It is therefore possible to conclude that vocabulary continues to play a key role even later on, throughout the fourth and fifth year of life, in the development of early Italian (as far as morphological development is concerned, at least).

14.2.3.2. Syntactic Productivity, Vocabulary and Age

As predicted vocabulary is a better predictor of syntactic productivity with the **nonce verb** than age (table 14.6b; fig. 14.10). Productivity with the nonce verb can be thought of as the measure of the extent to which children can retrieve (and rely on) fully-schematic units. This is because productive answers under the nonce verb condition suggest that children can apply a rule/pattern (the declarative transitive schema *(S)+V<O*) to an item (*bodare*) never encountered in that specific morpho-syntactic environment. Since vocabulary can be thought of as a measure of the opportunities a child has had to learn lexically-specific patterns from which s/he could have inferred more schematic units, it is consistent with a UBA that vocabulary is a better predictor of syntactic productivity than age[87].

Contrary to predictions however, age appears to be a better predictor of syntactic productivity with the **familiar** verb than vocabulary (table 14.6b; fig. 14.10). This may suggest that age has a bigger role in syntactic productivity than in morphological productivity.

Productivity with the two types of verbs, it has been argued (11.5.1), may be construed as being underpinned by two different types of linguistic knowledge. Productivity with the familiar verb can be accounted for in terms of lexically-specific units. Conversely, productivity with the nonce verb is a function of a more schematic representation. On the basis of such an assumption, which

[87] See further discussion in 16.2.2.

appears to be borne out by the analysis in section 14.1.3, I will indulge in some speculation as to the role of age in syntactic productivity with the familiar verb. More specifically, I shall speculate (section 14.2.3.3) that such an effect could be explained in terms of Theory of Mind (ToM, henceforth; Harris, 2006; Slaughter, 2015) development. Afterwards, section 14.2.3.4 will briefly draw some conclusions on the relationship between age, vocabulary and syntactic productivity.

14.2.3.3. The effect of Age on syntactic productivity with the familiar verb

The classification tree in fig. 14.10 (node 2) clearly shows that it was the youngest children that were less productive with the familiar verb. Only 62% of children as old as thirty-nine months or younger (eleven two-year-olds and two three-year-olds) provided productive answers, as opposed to 97% (twenty-nine out of thirty) of older children who were productive with *lanciare* "to throw".

I would argue that such an age-effect is caused by cognitive-social development factors that partly neutralised the design's control for DO null realisation.

The design attempted to prevent DO null realisation by creating the pragmatic conditions for the patient to have a high value of informativeness. Children interacted with *Silvia*, a plush mole who was both narcoleptic (if she is asleep she is aware of neither the nature of the action nor who was acted upon) and blind (the animal cannot be aware of the physical presence of the patient)[88].

Such forms of control rely on:

a) The child's ability to role-play and pretend that *Silvia the Mole* and E., who moves and voices her, are two distinct [+ANIMATE] entities.

b) The child's ability to infer that *Silvia* on the one side, and child and E. on the other, do not share the same Knowledge Access (Slaughter, 2015; Peterson, Wellman & Slaughter, 2012) as to who did what to whom.

Regarding (b), the beliefs people have about specific situations, more often than not, are a function of their Knowledge Access to those situations. For instance, Mark puts his sweet in a blue box and leaves the room in which the box is. Upon his return to the room, he will look for the sweet in that same blue box, even

[88] Refer back to 11.5.4.2.

though Sarah moved it into a red box during his absence. Mark's action (looking for the sweet in the blue box) is driven by his (false) belief that the sweet is still in the blue box. Such a false belief is a by-product of his defective Knowledge Access, i.e. the fact that he put the sweet in the blue box, but he did not see Sarah move it into the red one.

Slaughter (2015) reports that, when twenty-five-month-olds watch a video depicting the scene just described, they anticipate Mark's movement and immediately look at the blue box, where he will look for the sweet. Conversely, six-year-old autistic children, whose ToM abilities are typically underdeveloped, do not look at the blue box (they do not anticipate Mark's next move). This suggests that normally-developed children understand that Mark still believes (wrongly) that the sweet is in the blue box (because that is where he put it), and he will act accordingly. Yet, when three-year-olds are asked to report to experimenters where Mark will look for his sweet, they answer that he will look for it in the red box and only manage to answer correctly (the blue box) one year later (Harris, 2006; Peterson et al., 2012; Wilde-Astington & Edward, 2010). Airenti (2015) makes sense of such contradictory results[89] by arguing that young children have some implicit ability for this kind of mindreading (i.e. some implicit ToM), but that explicit knowledge emerges only later[90].

As for point (a), Kavanaugh (2006, pp. 153-155) reports that two-year-olds are capable of fairly complex role-plays, such as both passive and active agency and sometimes even of more articulated shared pretend play. The fact that children were compliant with the task and happily answered Silvia's question (*What did Agent do?)* appears to imply that they were able to role-play and tell E. and *Silvia* apart.

It also means that they understood that *Silvia* could not be aware of what went on (she had a different Knowledge Access from the child), because she was

[89] i.e. the disparity between the twenty-five-month olds' attraction to the blue box (the correct answer) and the three-year-olds' inability to report this correct answer.

[90] I would further argue that such outcomes appear to be consistent with Munakata et al.'s (1997) hypothesis that different tasks (looking vs. telling) require different degrees of representational strength (implicit/weak vs. explicit/strong).

asleep while the game was played. That is, they realised that *Silvia* could not know anything about the specific game played by the two toys.

When production is elicited, though, only the Agent, the Patient and the little catapult used to perform the action are on the scene. The question activates the Agent-toy and makes it part of the shared knowledge between the child and Silvia. The DO is left linguistically inactive. However, its status of activation also depends on extra-linguistic factors, such as physical presence (see Tedeschi, 2007).

It is by virtue of Silvia's lack of visual access that the patient assumes a high value of informativeness, as her blindness prevents it from being part of the child and mole's shared knowledge (it neutralises the role of physical presence). Crucially though, for this form of control to work, children must somehow be able to put themselves in the shoes of someone who is blind and infer that her blindness gives her a different Knowledge Access to the scene. This might be cognitively demanding for two-year-olds who are not visually impaired. Indeed, such an inference is likely to rely on that kind of explicit meta-analysis that starts to develop during the fourth or fifth year of life. If the child fails to infer that Silvia cannot see, the patient is the only candidate as the landmark of the agent's action and its physical presence makes it accessible to both Silvia and the child. Because of that, it assumes a low value of informativeness and becomes a potential candidate for null realisation.

Assuming that such a line of reasoning is plausible, the argument put forward presents three main questions:

a) If inferring the blindness of the mole depends on ToM development, why would most of the two-year-olds (62%) somehow manage to be successful (i.e. linguistically productive)?

b) If such an *explicit* understanding of others' beliefs starts developing during the fifth year of life (see Harris, 2006; Peterson et al., 2012), why would three-year-olds behave so differently (88% of realised DOs with the familiar verb) from two-year-olds?

c) If ToM-related factors were behind syntactic productivity with the familiar verb, why couldn't (lack of) productivity with the nonce verb be explainable by those same factors?

I shall discuss these questions, accordingly.

a) Throughout the familiarisation phase, the children were repeatedly told that Silvia could not see and that she needed "to be told things". There were many explicit explanations and demonstrations that her Knowledge Access was different. For instance, when awake she would ask a child *What are you doing?* If the child's answer was inappropriate (e.g. s/he just answered by saying *this* and showed the mole what s/he was doing), Silvia would point out that she could not see and E. would encourage the child to be more specific (*I'm putting sand into the truck*, *I'm combing my doll's hair*). Indeed, the majority of children provided appropriate answers by the end of the familiarisation. Harris (2006) notices that children's performance in ToM-related tasks improves across all age-groups when they are provided with explicit guidance (verbal or visual) as to what others' beliefs are. Hence, it might be that younger children who were successful benefited the most from the familiarisation phase.

b) Tedeschi's (2007) results suggest that three-year-olds (3;6-4;1) show some sensitivity to the discourse pragmatics of questions: they dropped more objects after specific (35%) than after generic (11%) questions. Hence, it is arguable that, given this emerging sensitivity, they did not need to fully utilise all the control conditions designed. Instead, they just needed to attend to the question asked by the Mole, which assigned a high value of informativeness to the patient.

c) Clearly, it is not possible to rule out that the same ToM-related factors played no role in productivity with the nonce verb. Indeed, if such factors played a role in production with the familiar verb, it would be unreasonable to posit that they had no role in production with the nonce verb. Yet, the children's poor performance with the latter cannot be explained away in light of such factors.

Firstly, most (62%) of the children as old as thirty-nine months or younger were productive with the familiar verb, but only one of them (17%) was productive with the nonce verb. If pragmatic and ancillary factors were the only reasons behind productivity with the nonce verb, then performance on the two verbs, I would argue, should be (statistically) similar. For both two (z=-2.16, p=.031) and three (z=-2.21, p=.027) year-olds the proportion of productive participants was smaller in the nonce verb condition than in the familiar verb condition (table 14.1b; fig. 14.2, node 2). Furthermore, three-year-olds were clearly mature enough to use one or all of the design manipulations to infer the low informativeness value of the DO, as 88% of them provided productive answers with the familiar verb. Hence, their poor performance with the nonce verb (40% of productive participants) is not a by-product of pragmatic factors, but reflects differences in (linguistic) representational strength. This interpretation is further borne out by the fact that all children's age-groups performed statistically similarly to adults with the familiar verb, but only four-year-olds performed in an adultlike manner with the nonce verb (table 14.2).

14.2.3.4. Final remarks on Age, Vocabulary and Syntactic Productivity

A further argument that suggests that results pertaining to the nonce verb cannot be explained away in terms of ToM development is the fact that vocabulary, and not age, is the best predictor of syntactic productivity with *bodare*. Indeed, if ToM-related factors were the primary reasons behind the poor performance with the nonce verb, why would vocabulary (a measure of linguistic experience) be a better predictor of syntactic productivity with the nonce verb than age (a variable more likely to be tightly linked to children's general cognitive and social development)?

The children were overwhelmingly productive with the familiar verb (84% of productive participants), but not with the nonce verb (47% of productive participants; fig. 14.10, node 1, p<.001). Node 2 in fig 14.10 shows that the age effect on productivity with the **familiar verb** separates the youngest children, who were **mostly** (62%) productive, from older children, who were **overwhelmingly** (97%) productive. Node 3 (fig. 14.10), instead, splits a group of children (vocabulary score ≤ 67) who were **not productive at all** (0% of productive participants) and another group (vocabulary score > 67) who were **mostly** (60%) productive.

My hypothesis is that those unproductive children can combine the familiar verb *lanciare* "to throw" with a DO, but because the design did not perfectly control for the pragmatic conditions behind null argument realisation in early language, some of them simply dropped the DO. Since attending (inferring) all the pragmatic conditions of the study partly depends on ToM development, younger children were less productive than older children. Yet, most children in both groups *were* syntactically productive with the familiar verb.

Conversely, participants with a poor vocabulary score (≤67) were not productive at all with the nonce verb. One could regard such results as evidence that children need to learn a critical mass of examples (vocabulary > 67) in order to start developing constructional schemas. Such an interpretation would be consistent with both the results of this study and the hypothesis that the entrenchment of more schematic constructions is solidly linked to quantity (and quality) of linguistic experience. Importantly, this is not to say that cognitive development plays no role in the development of schematic constructions. Indeed, in order to develop such units, children must draw generalisations based on form-function mapping. In order to map a form onto a meaning, children must have developed the necessary cognitive abilities to understand such meaning (see Bates & MacWhinney, 1987). Furthermore, more general, non-linguistic abilities, such as categorisation, schematisation and joint attention must be adapted to (and for) the task of language learning and use.

I am therefore tempted to conclude that vocabulary is a better predictor of syntactic productivity on the basis of the following observations:

a) vocabulary is indeed a better predictor of syntactic productivity with the nonce verb than age (table 14.6b; fig. 14.110);

b) productivity with the nonce verb is a function of the entrenchment of schematic units;

c) two and three-year-olds performed in an adultlike manner with the familiar verb, but not with the nonce verb (table 14.2);

d) two and three-year-olds, but not adults and four-year-olds, performed better ($p < .05$) with the familiar verb than with the nonce verb (table 14.1b);

e) the effect of age can be explained in terms of ToM development.

Such a conclusion however, can have only provisional value as further studies are needed. Such studies will have to take into account children's ToM development while designing ways to control for the pragmatic conditions behind null argument realisation.

15.

CONCLUSION

Now that the reader has an understanding of the various analyses of the outcomes of this experimental study, the research questions posed in chapter 10 are answered in the following paragraphs.

a) **Is there any evidence that Italian speaking children can rely on (or retrieve) fully schematic constructions in order to be *morphologically* productive?**

According to the statistical analysis, Italian-speaking children aged 2;02-3;00 (M=2;08), if considered as a group, showed ability to inflect in the past participle both familiar and newly learnt (the nonce verb) Conjugation I material with similar mastery (z= -1.03, p>.05, table 14.1). This suggests that a significant proportion of them can retrieve the schema *PROCESSato* to productively use newly learnt verbs. Importantly, the fact that some (33%) of them were able to inflect a nonce verb in its regular past participle may be regarded as evidence that those children were able to parse the experimental stimuli *and* they had developed a network of morphological constructions that links various Conjugation I inflections. As shall be discussed in more detail in 16.1.4, this appears to suggest that those (productive) children were somehow able to categorise both the experimental stimuli and the schema *PROCESSato* as instantiations of a fully schematic unit *PROCESS-INFLECTION*. This in turn implies that, at least to a certain extent, those two-year-olds were able to evoke a fully schematic construction (*PROCESS-INFLECTION*) in order to bring their knowledge of the schema *PROCESSato* to the experimental task.

Thus, the statistically (p>.05) similar performance across verb_familiarity conditions may be interpreted as indicating that, **as a group**, two-year-olds were able to evoke schematic constructions (productivity with the nonce verb) as easily as they evoked lexically-specific units (productivity with the familiar verb).

Importantly though, such schematic (morphological) knowledge does not yet appear to be fully entrenched as adults performed significantly (p<.05) better than two-year-olds. It is only at the age of three years that children appear to have developed an adultlike morphological productivity.

b) **Is there any evidence that Italian speaking children can rely on (or retrieve) fully schematic constructions in order to be *syntactically* productive?**

Two-year-olds and three-year-olds, but not adults and four-year-olds, performed better with the familiar verb than with the nonce verb (table 14.1b; fig 14.2). Hence, it is not possible to say that Italian-speaking children aged two to three years have developed the fully schematic transitive constructional schema. Nevertheless, the pattern that emerges shows that the development of fully schematic constructions is not an in/out distinction; rather, it is a matter of degree. As the years go by, the difference in productivity between the two verbs decreases, until it disappears during adulthood (fig. 13.3). The increase in productivity with the nonce verb (fig. 13.3 and 14.8) as participants grow older can be interpreted as a function of the piecemeal entrenchment of the Italian transitive construction. As the constructional schema becomes more and more entrenched, children find it increasingly easier to retrieve (and rely on) such a schema; this yields an increasing number of productive participants as a function of age_group. Importantly, by the age of four, children perform in an adultlike manner with both types of verbs and similarly across verb_familiarity conditions, which may be seen as suggesting that they have acquired the declarative transitive constructional schema.

c) **Given previous findings regarding both Italian and English-speaking children, would vocabulary be a better predictor of morpho-syntactic productivity than age?**

Vocabulary is a better predictor of **morphological productivity** than age with both the familiar verb and nonce verb. This is not only in line with previous findings regarding English-speaking children's over-regularisations of irregular participles (Marchman & Bates, 1994), but it also indicates that vocabulary development continues to play a key role throughout the fourth and fifth year of life.

Syntactically, vocabulary (a measure of children's linguistic experience) is still a better predictor of productivity than age with the nonce verb (table 14.6b; fig 14.10). This finding is intriguing as it appears to suggest that the entrenchment of fully-schematic units can be regarded as a function of vocabulary development.

However, age is a better predictor of syntactic productivity with the familiar verb than vocabulary. Although such an outcome could be accounted for in terms of the pragmatic conditions behind child null argument realisation, it nevertheless calls for further research to establish whether such an interpretation holds correct. Hence, even though it could still be argued that vocabulary is a better predictor of syntactic productivity than age, certain conclusions cannot be drawn from the results of this study.

Page intentially left balnk

Part IV:

Conclusions

Page intentionally left blank

16.

GENERAL DISCUSSION

Now that the results and analysis of the both experimental and longitudinal studies have been presented, the current chapter further discusses the findings in terms of morphological (16.1) and syntactic (16.2) development in a unified manner.

16.1. MORPHOLOGICAL DEVELOPMENT

Both the longitudinal and experimental studies suggest that productivity with Italian regular morphology starts to show up as early as during the third year of life. About 2% of Roberto's target sentences are Soft_Constructional_Fails at the word level. These are sentences containing words attested in the main corpus in only one morphological form (e.g. M.SG) that does not match exactly the form the child uses (M.PL). Thus, Roberto appears to be capable of inflecting a regular noun at, for instance, its M.PL form, even when the input does not appear to have provided him with information as to which is the exact form the word takes when plural. Similarly, although their (group) performance was far from perfect (33% of productive participants), two-year-olds who participated in the experimental study, were (statistically) equally morphologically productive across verb_familiarity conditions ($z=-1.03$, $p>.05$).

16.1.1. MORPHOLOGY BEFORE SYNTAX

Consistent with previous research regarding English-speaking children (e.g. Akhtar & Tomasello, 1997), the results indicate that some aspects of morphology get off the ground earlier than syntax. Indeed, when it came to production with the nonce verb, each age group was more productive with morphology than syntax (fig. 16.1, overleaf). This is most evident in the results pertaining to three-year-olds, whose proportion of syntactically productive participants with *bodare* was lower than the proportion of syntactically productive adults, yet they were morphologically as productive as adults with both the familiar verb and the nonce verb (refer back to table 14.2).

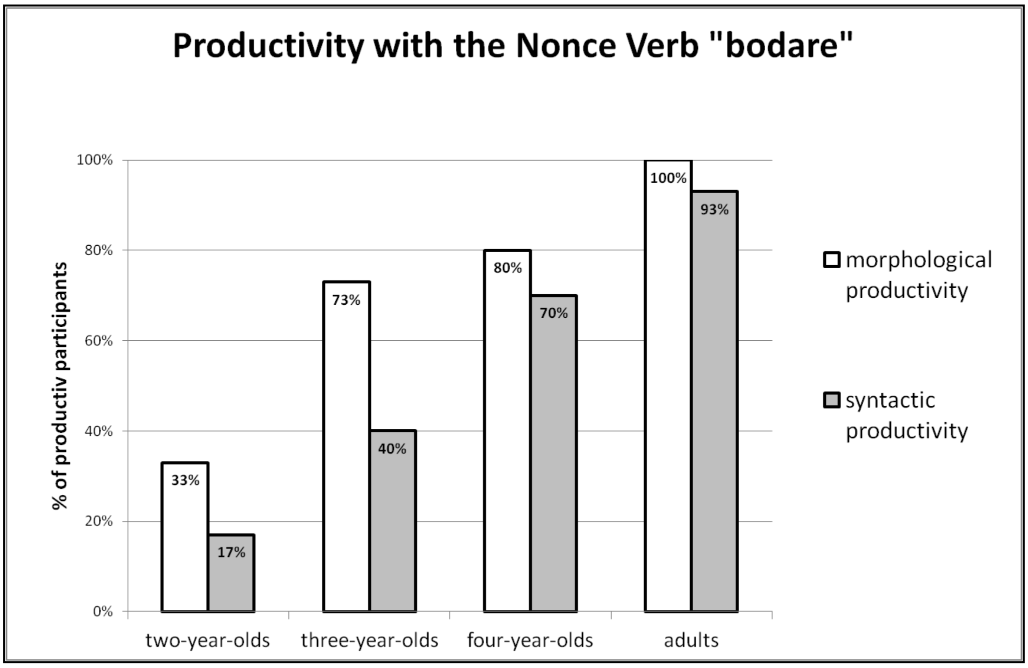


Figure 16.1: The proportion of productive participants with the nonce verb in each age-group.

The interaction of three factors is likely to be behind such a developmental sequence:

a) the *passato prossimo* schema represents – in Bates and MacWhinney's (1987) terms – more of a local cue than the transitive schema;

b) attending to the strings that instantiate the *passato prossimo* is likely to require less working memory than attending to the strings that instantiate the transitive construction;

c) the various instantiations of the *passato prossimo* construction share a recurring concrete element (-*t*-) that consistently maps onto a specific meaning (past), which is likely to facilitate analogy (and hence schematisation).

16.1.1.1. Local vs. global cues

It could be argued that, in both Italian and English, verb morphology is a local cue (at the word level), whereas syntax is more of a global cue (at the clause or sentence level). Since local cues are easier to acquire than global ones, children become productive with morphology earlier than they do with syntax.

However, the *passato prossimo* construction appears to lie more towards the topological end of the local-global spectrum as in order to deliver the correct form, children must consider both the main verb and the auxiliary verb and their relationship; ***ha*** *bod**ato*** “**has** nonce**.**verb**ed**”.

As for the naturalistic study, Roberto’s dataset attests nine overgeneralisations of the regular Conjugation III past participle *-i-to* “TV(conj.III)-PTCP” (refer back to table 1.11) to the irregular Conjugation II *rompere* “to break” (**romp-i-to* “break(root)-TV(conj.III)-PTCP” instead of *rotto* “broken”). Importantly, he does so by using various forms of the auxiliary: PRS.1.SG, PRS.3.SG and PRS.3.PL. Thus, if the results of both studies are considered together, it appears that at least some two-year-olds may have acquired the multi-word schema *AVERE PROCESS-TV-t-GENDER.NUMBER* “HAVE PROCESS-TV-ptcp-GENDER.NUMBER”[91].

Such a schema presents a notable degree of abstraction (four slots), which is hardly different from the one on which adults supposedly rely. Furthermore, its “multi-word nature” collocates it more on the topological, rather than on the local end of the spectrum. Consequently, one may argue that the earlier emergence of *passato prossimo* constructions cannot be fully accounted for in terms of local vs. global cues.

Yet, as Abbot-Smith and Serratrice (2013) point out, the “local vs. topological” distinction is a matter of degree[92]; the more global the cue, the heavier the burden placed on children’s processing. The transitive *(S)+V<O* schema might indeed be considered more of a global “phenomenon” than the *passato prossimo* schema in one important respect. The optional subject adds a further element to be considered, which clearly increases the number of elements to be put in relation to each other (hence, it further pushes the construction towards the global/topological end of the continuum). The *passato prossimo* schema

[91] One may argue that the slot representing generalisations across various TVs is a “grown-up” generalisation. Children might have simply learnt to add the suffix *–at-GENDER.NUMBER* to the verb root. However, this is less likely to be the case for Roberto, as he generalises a regular Conjugation III past participle to an irregular Conjugation II verb. This is a sensible choice, as some Conjugation II no-*root-change* verbs form their participle using the Conjugation III TV –*i* (refer back to 1.1.2.3). Thus, Roberto’s production shows some emerging sensitivity to the role TVs play in determining the inflectional behaviour of verbs.

[92] For instance, as noted in 8.4.5.1, case-marking, word-order and subject-verb agreement posit themselves on a continuum in which case-marking is “fully” local, word order is more global than case-marking and subject-verb agreement is even more global than word-order.

requires children to generalise across two-element sequences, while the transitive schema may require that children process three-element sequences.

16.1.1.2. Working memory

A further aspect that needs to be considered is working memory.

The transitive schema could potentially be instantiated by only two elements (V<O), as the subject is frequently omitted. Thus, both schemas are likely to present two-element instantiations (AUX<V and V<O). However, subjectless VO patterns are likely to include more words than the instantiations of the *passato prossimo*. The latter schema can only be instantiated by two-word strings (AUX<V). Conversely, both the verb and the DO may be instantiated by fairly long strings (e.g. *I* ***[will be eating]*** **[*your favourite chocolate cake]***). Longer strings may require more working memory than shorter strings. This is likely to put a heavier burden on children's processing faculties: in order to be able to generalise across those longer strings, children have to keep in mind more elements (and their attention span must therefore be longer).

16.1.1.3. Children's analogy skills

Children's analogy skills are also an important factor that could account for the early entrenchment of the *passato prossimo* schema.

It has been previously pointed out (8.1) that the co-occurrence of functional and perceptual identity across elements of the structures being analogised facilitates functional analogy. This is because identical elements are constantly associated with a fixed function (Tomasello, 2003; 2006b). This is borne out by evidence from both linguistic (Childers & Tomasello, 2001) and non-linguistic (Gentner & Markman, 1995, 1997; Gentner & Medina, 1998) studies.

Perceptual similarity is hardly a factor in the various instantiations of the transitive schema. Conversely, the morpheme *–t-*, which marks the past participle, does provide a small, yet significant anchor of perceptual similarity (fig. 16.2, in blue). Such an anchor of perceptual similarity is likely to facilitate functional analogy across different instantiations of the *passato prossimo* schema, as it provides an element of phonological sameness (*-t-*) that constantly maps onto a specific meaning (past). That is, the affix *–t-* makes form-meaning mapping very transparent. The result is that the schema is acquired quite early on in development.

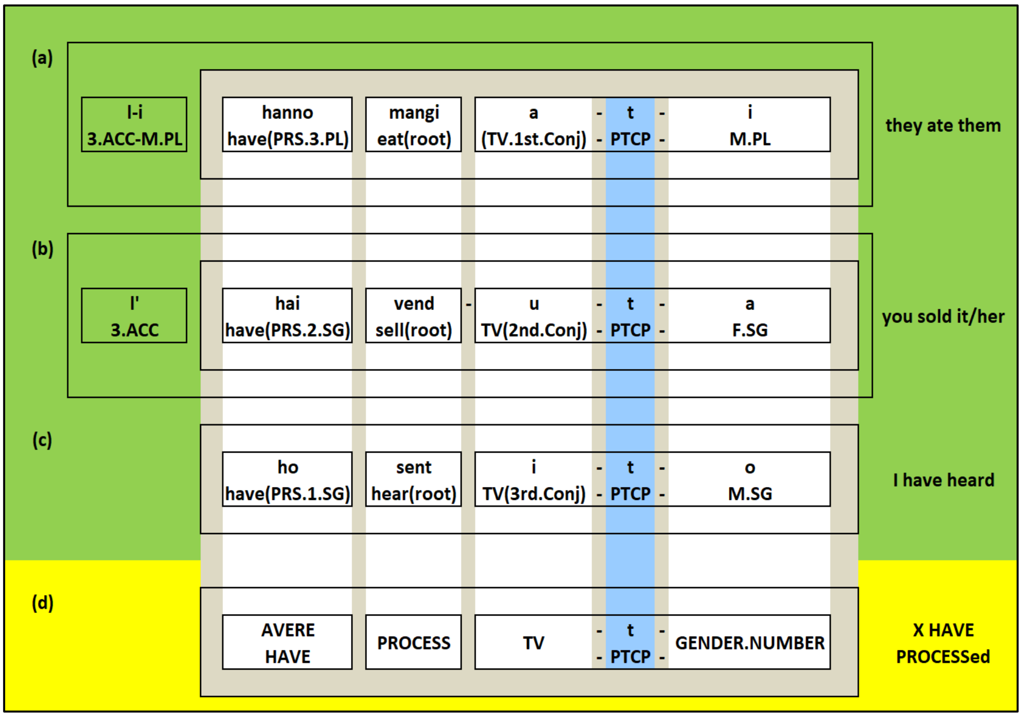


Figure 16.2: Abstracting the *passato prossimo* schema (in the grey square). Concrete strings are included in the green strip, whereas the schema appears in the yellow strip. The blue highlighting indicates fixed recurring material that consistently maps onto a fixed meaning (past). Slot formation is highlighted in white.

16.1.1.4. Conclusion

Overall then, it may be argued that the earlier acquisition of the *passato prossimo* schema with respect to the transitive construction is brought about by the fact that acquiring the former is cognitively less demanding than acquiring the latter.

The *passato prossimo* schema:

a) may require that fewer elements be put in relation to each other (it is more of a local cue);

b) is more likely to be instantiated by shorter strings (less working memory needed);

c) presents a recurring element (-*t*-) that provides an anchor of perceptual and semantic sameness that is likely to facilitate analogy.

16.1.2. CROSS-LINGUISTIC DIFFERENCES

If the results of this research are compared with Akhtar and Tomasello's (1997) findings regarding the acquisition of the schema *PROCESSed*, Italian-speaking children appear to start developing productivity with past participles earlier than their English-speaking peers. Indeed, the proportion of morphologically productive two-year-olds is higher for Italian (33%) than English (15%). Such findings can be accounted for in terms of language-specific factors.

As noted earlier (14.1.4), form-function mapping is supposedly easier and more reliable for the Italian Conjugation I past participle than for its English counterpart *PROCESSed*. Indeed, the Conjugation I past participle represents an ideal candidate for schema abstraction:

a) Conjugation I is the largest inflectional class and the overwhelming majority (70%) of Italian verbs belong to it (token and type frequency);

b) with the exception of four verbs and their derivatives, Conjugation I verbs are all regular. In fact, three of these four irregular Conjugation I verbs (*andare* "to go", *stare* "to stay" and *dare* "to give") add the regular affix *–ato* to their root (e.g. *and-a-to* "go-TV(conj.I)-PTCP") to form the past participle. Hence, there are virtually no forms that enter into competition with the regular past participle once it is established that the verb belongs to Conjugation I;

c) the inflection *–ato* is bi-syllabic and stressed (perceptual saliency).

The interaction of regularity, phonological saliency and token and type frequency facilitates schematisation, as functional analogies are easier to draw across regular and frequent patterns. As a result, the schema *PROCESSato* starts being acquired quite early on in development.

Similar arguments in terms of transparency of form-function mapping can be made about the Italian regular nominal morphology. Italian has four regular gender and number markers (*-o* "M.SG", *-i* "M.PL", *-a* "F.SG" and *–e* "F.PL") which apply to 75% of nouns (see Tartaglione, 1997; and section 1.1.1), articles, all adjectives that inflect for both gender and number, 3.ACC clitic pronouns and past participles. Thus, the frequent and regular recurring forms in fig. 16.3a should facilitate the formation of the schema in fig. 16.3b. Similarly, the forms in fig. 16.3c should facilitate the development of fig. 16.3d. The formation of the two lexically-bound schemas lays the foundations for the abstraction of the fully schematic unit in fig. 16.3e.

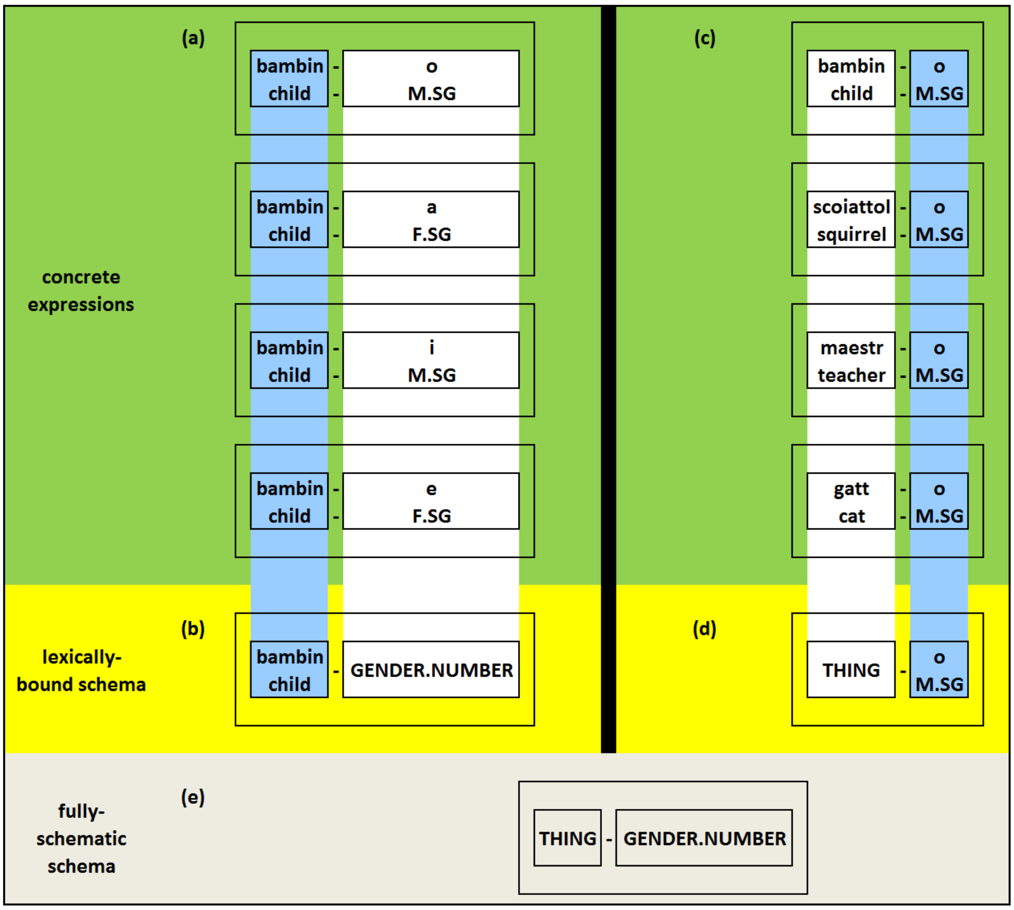


Figure 16.3: abstracting morphological schemas in Italian. Concrete expressions are in the green strip and lexically-bound schemas in the yellow strip. The fully schematic unit abstracted from the two lexically-specific schemas is in the grey strip (e). Shared lexical material is in blue and relationships of elaboration (i.e. slot formation) are highlighted in white.

As noted in the previous section, Roberto overgeneralises the regular Conjugation III past participle to the irregular Conjugation II *rompere* "to break". This appears to indicate that (some) Italian-speaking two-year-olds might indeed master regular past participles in a "cross-conjugational" way.

Most (about 90%) of Conjugation III verbs are regular and some Conjugation II *no-root-change verbs* form the past participle in the same way (i.e. by using the same TV) as regular Conjugation III verbs (refer back to 1.1.2.3 and table 1.3). Hence, one may speculate that, initially, the schema *PROCESS-TV-t-GENDER.NUMBER* is a generalisation between Conjugation I (fig. 16.2a) and

Conjugation III (fig. 16.2c) past participles to which regular conjugation II past participles (fig. 16.2b) are assimilated only later. Although this does not seem implausible, empirical research (both in naturalistic and experimental settings) is needed to investigate whether this holds true[93].

Nevertheless, the very fact that irregularities are indeed present (and frequent in the case of Conjugation II) in the Italian verbal system suggests that transparent form-function mapping is unlikely to be the only factor at play.

Devescovi et al. (2005) and Caselli et al. (1999) provide another piece of explanation. As briefly discussed earlier (14.1.4), both studies found that when matched for age and vocabulary, Italian-speaking children show a more mature morphological behaviour than their English-speaking peers. They propose that such a different pace of morphological acquisition is a by-product of children's sensitivity to the characteristics of their ambient language. Italian children are morphologically more advanced than English children because morphology plays a bigger role in Italian than English (Devescovi et al., 2005). For instance, 70% of Italian sentences are subjectless (Lorusso et al., 2005). This means that in order to understand who the entity in the foreground of the VP[94] is, children need to attend to verbal inflections.

Devescovi et al. (2005) calculated different measures of the extent to which children's sentences (*Kitty sleeping*) differed from their adultlike targets (*Kitty is sleeping*) for both Italian and English-speaking children. Their analysis yielded two main findings. Firstly, as previously mentioned, Italian-speaking children, on the whole, produced utterances that were morphologically more complex than the utterances produced by their English-speaking peers. Secondly, however, the gap between attempted and target utterances was similar across the two language groups. Hence, it is not that Italian-speaking children are linguistically more advanced than English-speaking children *tout court*. The former are more advanced in those aspects for which their ambient language provides a much richer and communicatively more relevant input. In other words,

[93] I am not aware of any research that explicitly investigated such an issue.
[94] In Langacker's (1987, 1991) terms.

> “The pace and complexity of development varies with complexity in the child’s input”.
>
> (Devescovi et al., 2005, p. 782)

Put simply, language is used to communicate and children more carefully attend to those aspects of their ambient language that ensure a more successful communication.

Thus, the more advanced morphological behaviour of Italian-speaking children with respect to English-speaking ones is likely to be a by-product of language-specific factors. Morphology is likely to be more relevant for communicative purposes in Italian than English. Additionally, the Italian inflectional system is highly regular, frequent and phonologically salient, which makes form-function mapping a relatively easy task.

16.1.3. MORPHOLOGICAL DEVELOPMENT AND VOCABULARY GROWTH

Despite the fact that most English-speaking two-year-olds could not be productive with the schema *PROCESSed* in Akhtar and Tomasello’s (1997) study, Marchman and Bates (1994) found that at least some English-speaking two-year-olds can be said to be productive with that same schema, as they occasionally over-regularise irregular past participles (**goed, *throwed*). Marchman and Bates (1994) found that the number of over-regularisations and vocabulary size (verbs only) were correlated ($r + .56$, $p < .00001$; refer back to 3.9.2). The results of the current research are consistent with their findings, as morphological productivity appears to be a function of children’s vocabulary size rather than their age. This is more evident from the analysis of the experimental study (table 14.4b; fig. 14.7). It is also interesting to note that Roberto, who can be said to be productive with regular nominal morphology and also overgeneralises regular past participles, has a vocabulary that is above the 75th percentile for 30-month-olds. Hence, his morphological productivity can be explained in light of his large vocabulary. These findings replicate a stream of studies led by Elizabeth Bates (Marchman & Bates, 1994; Bates & Goodman, 2001; Caselli et al., 1999; Devescovi et al., 2005) which showed that vocabulary is a better predictor of linguistic development than age. Importantly, the experimental study presented here enquired into an age range (02;02 to 05;00) that is much wider than the one Bates and colleagues investigated (0;10 to 2;08). Hence, it is possible to conclude that vocabulary continues to play an important role in morphological development throughout the pre-school years.

16.1.4. EARLY NETWORKS

The early emergence of both nominal (the longitudinal study) and verbal (the experimental study) morphological productivity can be interpreted as evidence that (at least some) two-year-olds have started developing networks of constructions that link morphological inflections (i.e. word-level schemas) with each other.

However, the inferences that can be drawn about the nature of such networks are not quite the same in terms of adultlike form-function mapping, as they are based on different types of data: naturalistic and experimental.

16.1.4.1. The experimental study (verbal morphology)

As for the experimental data, the fact that, when considered as a group, even two-year-olds were as (statistically) productive with the nonce verb as with the familiar verb suggests that a significant proportion (33%) of them could recognise the experimental stimulus as a *PROCESS* whose morphological behaviour assigned it to Conjugation I. This in turn implies that those children:

a) could parse the experimental stimuli as *bod-a* "nonce.verb-IMP.2.SG" and *bod-a-re* "nonce.verb-TV(conj.I)-INF";
b) had developed the schema *PROCESS-a-to* "PROCESS-TV(conj.I)-PTCP" and could link those forms to each other.

Overall, results could be interpreted as evidence that those two-year-olds have abstracted a superordinate schema (fig. 16.4a, overleaf) that subsumes at least the morphological inflections used and elicited in the experimental design (fig. 16.4b-d)[95]. Importantly, such links across constructions are evidence that those schemas are shaped in an adultlike manner in terms of form ([[/bod/]>[/a/]]) – meaning ([[BODARE]-[IMP.2.SG]]) mapping.

[95] But see note 91; the inclusion of the TV in fig. 16.4a simply indicates that the generalisations that can be inferred pertain to Conjugation I.

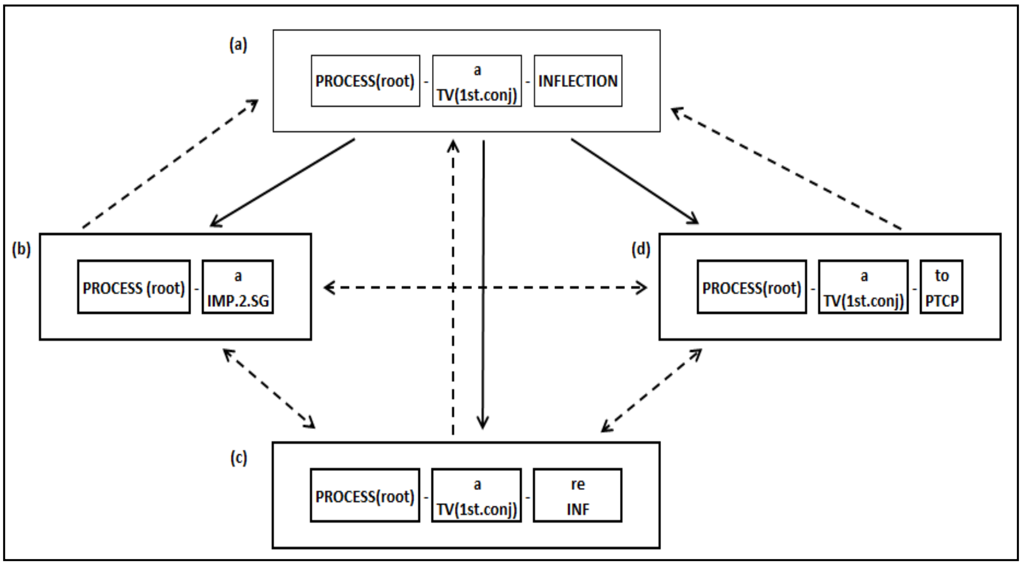


Figure 16.4: the putative network of morphological constructions that could be hypothesised as being developed by two-year-old Italian-speaking children. Dashed arrows indicate relationships of extension. Solid arrows indicate relationships of elaboration. Thickness of boxes indicates units' entrenchment.

Such a conclusion comes with two important caveats.

Firstly, although two-year-olds' group productivity was statistically similar across verb_familiarity conditions, fig. 13.2 clearly shows that most of them were unproductive with the nonce verbs (33% of productive children) and productive with the familiar verb (62%). It must be noted that two-year-olds' results on familiar and nonce verb are based on the answers provided by 13 and 12 children, respectively. Hence, the attested difference in productivity between the two verbs (-32%) might reach significance if more participants were recruited. Thus, such results call for replication studies in order to determine whether Italian-speaking children indeed start developing morphological productivity so early on in development (see also footnote 84).

Secondly, table 14.2 indicates that adults were significantly ($p<.05$, Bonferroni correction) more morphologically productive with the nonce verb than the youngest children. As noted earlier, the Fisher's Exact Test that compared adults and two-year-olds had an incredibly high odds ratio (Inf.) and CIs (3.62-Inf), which casts doubt as to whether such a result is, in fact, a Type I error (but it could also have been brought about by quasi-perfect separation). Be that as it may, the gap between adults and two-year-olds is undoubtedly much wider for the nonce verb (-67%) than for the familiar verb (-31%).

The overall picture can therefore be interpreted as follows. When group performance/productivity is analysed, two-year-olds appear to have developed a network of constructions that links together different Conjugation I inflectional schemas (fig. 16.4). The fact that at least some of them appear to have figured out that a nonce verb whose imperative is *bod-a* “nonce.verb-IMP.2.SG” (fig. 16.4b) takes *bodato* “nonce.verb-TV(conj.I)-PTCP” (fig. 16.4d) as past participle suggests that those productive children were able to categorise both of them as instantiations of a Conjugation I superordinate schema (fig 16.4a) that subsumes (at least) the Conjugation I imperative (fig. 16.4b), infinitive (fig. 16.4c) and past participle (fig. 16.4d). Being able to categorise 16.4b-d as instantiations of 16.4a is a cognitive activity and the more frequently such a cognitive activity is undertaken, the easier it becomes to engage in again. The entrenchment of the schema in fig. 16.4a is a function of the frequency with which such a cognitive activity is undertaken. Importantly, the frequency of such a cognitive activity is in turn a function of linguistic experience. As a result, two-year-olds’ schema (fig. 16.4a) is less entrenched than that of adults, because the latter had a lot more linguistic experience than young children[96]. Thus, adults perform better because their schema is more entrenched. Conversely, two-year-olds’ schematic units are less entrenched and hence less easily retrievable. This weaker representation yields a (group) productivity that is not quite adultlike. An implication of the above interpretation of the results pertaining to morphological development is that the acquisition of the schema *PROCESSato* (and hence linguistic productivity) is a matter of degree. Although the two-year-olds’ schema appears to be entrenched enough to allow a performance that is statistically similar across verb_familiarity conditions, it is still (substantially) less entrenched than the schema on which adults supposedly rely.

Finally, it has to be pointed out that the results of this study seem to indicate that at least some Italian-speaking children become morphologically productive earlier than hypothesised in previous studies, which showed that neither Italian (Pizzuto & Caselli, 1992, 1993, 1994; Leonard et al., 2002) nor Spanish (Aguado-Orea, 2004) speaking children younger than three years of age have an

[96] In fig. 16.4, the weakly entrenched status of the superordinate schema is represented by the fact that 16.4a is enclosed in a box that is thinner that the boxes in which 16.4b-d are enclosed.

across-the-board mastery of the inflectional system of their mother tongue[97]. However, the results are not inconsistent with those previous studies; whether the network hypothesised in fig. 16.4 includes (and therefore links) even other, less frequent Conjugation I inflections, or whether similar networks are developed for other conjugational classes are open questions that need further research.

What this experimental study can tell is that, as a group, even two-year-olds can use/activate the schema *PROCESSato* "PROCESSed" productively with both familiar and unfamiliar verbs. This may suggest that a significant proportion of them can parse a Conjugation I verb as root + inflection and link it to the other Conjugation I inflections used in the experimental design.

16.1.4.2. The longitudinal study (nominal morphology)

In section 8.2 it was claimed that Roberto appears to be morphologically productive with nominal morphology. At the same time however, when his target sentences have no concrete main corpus precedents which could account for gender-number agreement in lexically-specific terms, he produces more wrong (60%) than correct (40%) agreements (refer back to 8.4.4.2). These contrasting findings can be accounted for in terms of local (the formation of gender-number slots on single words) versus global (gender-number agreement) cues. Roberto has acquired the adultlike schema in fig. 16.3e, but since agreement spans across several words, he struggles to deliver correct gender-number agreement whenever he cannot rely on models that specify how to do that. Fig. 16.5 and fig. 16.6 report two attested schemas on the basis of which Roberto could produce correct gender-number agreement on a lexically-specific basis.

[97]For instance, Leonard et al.'s (2002) experimental study found dissociation in mastery of verbal inflections by Italian-speaking children. Two-year-olds showed error rates that peaked up to 30% (sd 41%) in production of PRS.3.PL This was not the case for PRS.1.SG (14% of incorrect answers, sd=29%). The implication is that children do not master plural and singular inflections with equal proficiency.

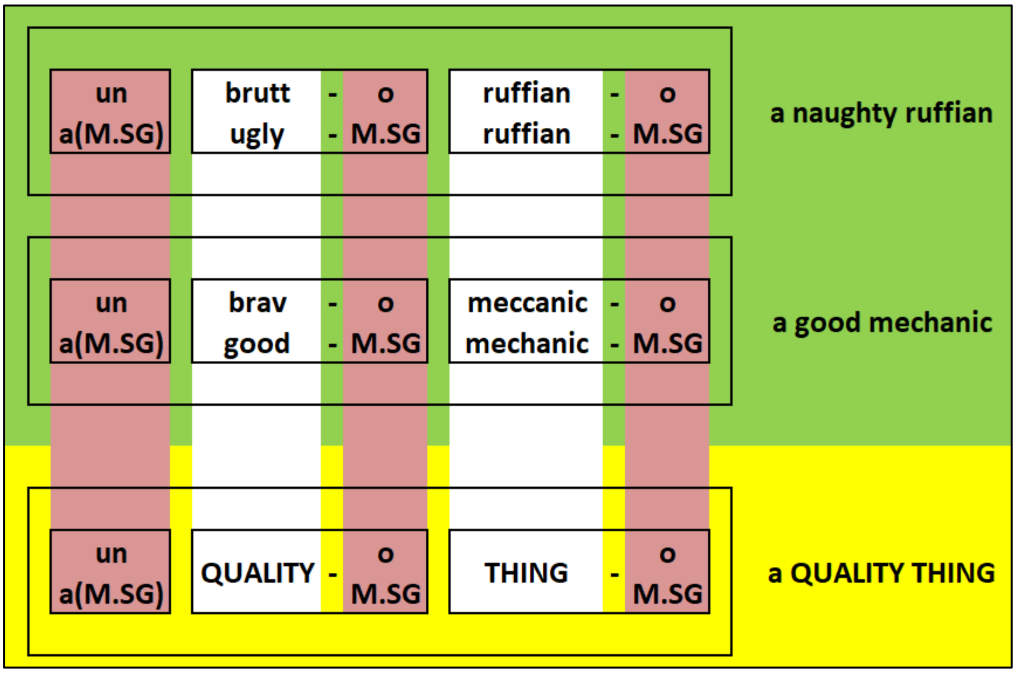


Figure 16.5: an attested schema that accounts for gender-number agreement across article, adjective and noun. Slot schematisation is in white. The concrete and recurring material that accounts for gender-number agreement is highlighted in red.

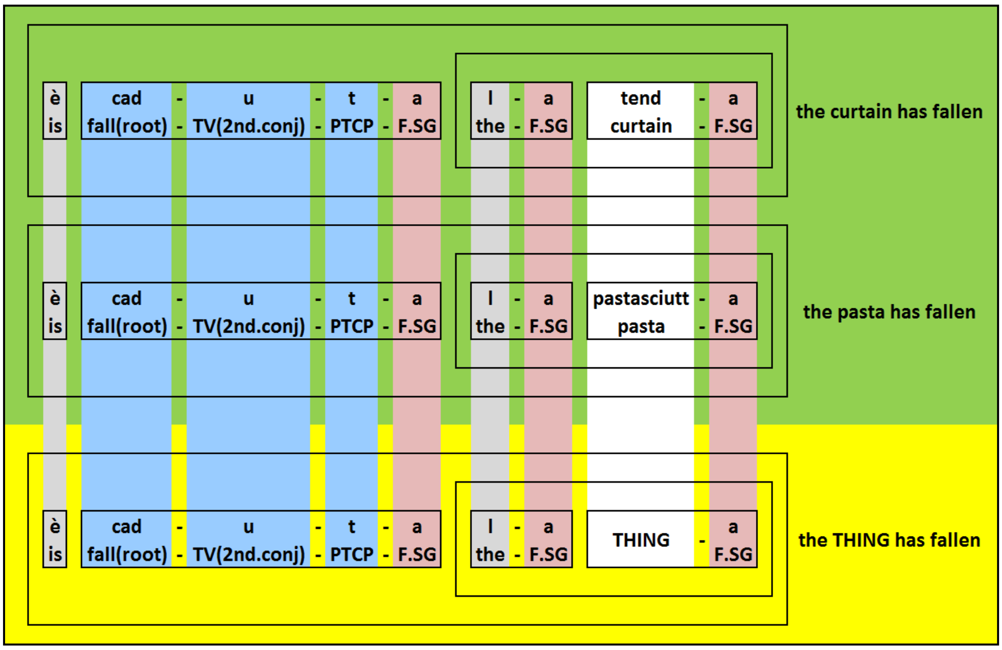


Figure 16.6: an attested schema that accounts for gender-number agreement across verb, past participle and subject-NP. Concrete and recurring lexical material which is not essential for agreement is in blue. Concrete and recurrent material that accounts for number agreement only is highlighted in grey. Finally, the red highlighting indicates concrete and recurring material that accounts for both gender and number agreement across elements. Slot schematisation is highlighted in white.

Such an interpretation in terms of the [+/- LOCAL] nature of cues is consistent with the previously (8.4.5.1) discussed results of Devescovi et al. (1998) pertaining to subject-verb agreement and thematic role assignment. When children deal with local cues (one-word stimuli; experiment 3), they act out verbal inflections correctly. When their knowledge of verbal inflections must be used to disambiguate thematic roles on a more global basis (i.e. they have to process NP1 and NP2, note the form taken by the verb and check which NP agrees with it), children simply ignore agreement (a global/topical cue) and follow local cues (namely, animacy; experiment 2).

Although this is indeed a possible explanation, the naturalistic nature of Roberto's data limits the types of inferences that can be made about the kinds of the generalisations he supposedly drew.

Referring back to sentence (53) and fig. 8.8 (section 8.2), the word *cerv-o* "deer-M.SG" is attested only in its M.SG form in the main corpus. Yet Roberto uses its root *cerv-* to fill the slot of the schema in fig. 8.8. Hence, he is able to parse *cerv-o* as root + M.SG and he also knows that its root can fill a slot ending with *–i* "M.PL" (*THING-i* "THING-m.pl"). This suggests that

a) he is capable of categorizing *cerv-o* "deer-M.SG" as an instance of the schema THING-o "THING-m.sg";
b) he has developed a network of constructions in which the schema *THING-o* "THING-m.sg" is linked to the schema *THING-i* "THING-m.pl".

However, points (a) and (b) above are disputable, as the very regular nature of the Italian morphological system could potentially allow learners to infer that the plural of *cerv-o* "deer-M.SG" is *cerv-i* "deer-M.PL" solely on the basis of phonological sensitivity. Indeed, Roberto's morphological productivity could potentially be explained by positing independent mini-networks based on phonological regularities that map the final vowel onto number but **not** gender information (fig. 16.7).

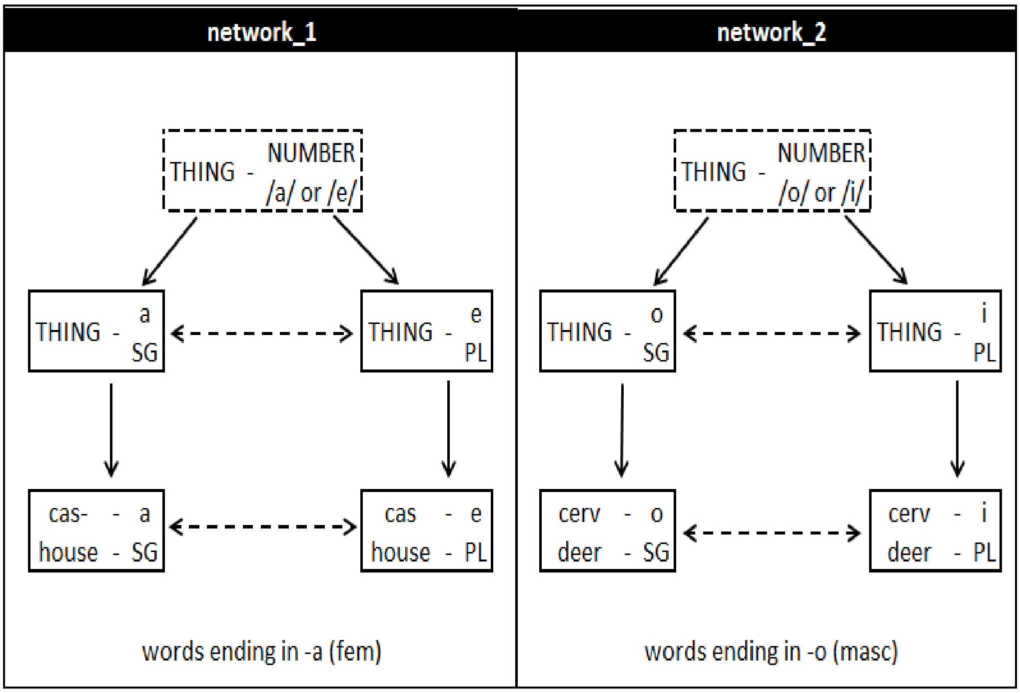


Figure 16.7: two independent networks of morphological constructions that account for gender-number flexibility. Network 1 can be paraphrased as "words that end in –*a* take plural form –*e*". Network_2 can be paraphrased as "words that end in –*o* take plural form in –*i*". Relationships of elaboration and extension are indicated by solid and dashed arrows, respectively. Thickness of boxes indicates units' entrenchment.

The following anecdote exemplifies the point I want to make. In the variety of Italian spoken by both E. and the participants, personal nouns (*Peter*, *Stephanie*, etc.) are often preceded by a determiner: *"la Stephanie* "the-F.SG Stephanie", *il Peter* "the(M.SG) Peter". Hence, the correct way to refer to E. is *il Luca* "the(M.SG) Luca". However, one of the children who participated in the experimental study constantly referred to E. as *la Luca* "the-F.SG Luca". This suggests that s/he generalised the schema *l-a THING-a* "the-F.SG THING-f.sg", without mapping the morpheme –*a* onto F.SG. The schema is a generalisation that appears to go as follows: *all names that end in –a are combined with "la"*.

What follows, is that without some form of experimental control, it is difficult to establish the extent to which Roberto's morphological productivity is a by-product of phonological sensitivity or, instead, whether he really mapped those final vowels onto both their grammatical functions. Undoubtedly however, he is capable of inferring that the plural of *cerv-o* "deer" is *cerv-i* "deer-M.PL", which appears to imply that he has developed a network of constructions that, at the very least, links singular and plural forms (fig. 16.7). Importantly, even networks like the one depicted in fig. 16.7 are evidence that Roberto has abstracted schemas and started drawing connections amongst them. What is not possible to establish, on the basis of naturalistic data, is the extent to which such constructions are mapped onto adultlike functions (i.e. grammatical gender).

16.2. SYNTACTIC DEVELOPMENT

The analysis of the outcomes of the experimental investigation indicates that four, but not two and three, year-olds have acquired the transitive construction and that they can retrieve that constructional schema to be as productive with newly learnt as with known material. The fact that the youngest children could perform in an adultlike manner with the familiar but not with the nonce verb has a twofold implication.

Firstly, it suggests that their syntactic competence is still mostly bound to a lexically-specific representation.

Secondly, when children can rely on well-rehearsed lexically-specific models their language is mostly adultlike.

The above interpretation of the experimental results appears to be borne out by the naturalistic enquiry into Roberto's multi-word sentences.

Firstly, about 90% of the syntactic patterns Roberto uses are lexically-specific schemas with a maximum of two slots. Furthermore, a thorough analysis of his language in terms of narrow and functional generalisations coherently accounts for both successful and unsuccessful derivations (only 2% of sentences are problematic). The implication is that the input children receive is rich and that their language can indeed be accounted for in terms of lexically-specific schemas inferred from the concrete strings that they have previously encountered.

Secondly, successful derivations, for which the method found frequent (putative) precedents (potentially encountered up to 20 times), are overwhelmingly (98%) grammatical sentences. Conversely, both Soft_Constructional_Fails and Hard_Constructional_Fails, whose identified (putative) models are less frequent (and hence represent less solid generalisations) and/or present more elaborative distance from their targets, are more likely ($p<.05$) to yield ungrammatical sentences (15% and 34% of ungrammatical sentences, respectively; see fig. 8.15 and table 8.4). These findings have two implications:

a) Relying on lexically-specific form-function mapping allows children to infer the right kinds of generalisations and therefore deliver mostly (98%) adultlike, well-formed sentences (fig. 8.15). At the same time, since lexically-specific schemas are productive units used to communicate, they might lead to over-generalisation errors (chapter 8.1).

b) When children push their form-function generalisations beyond well entrenched (lexically-specific) models, they are more likely to infer the wrong kinds of generalisations, thus delivering ill-formed (ungrammatical) sentences (section 8.4.2). Indeed, the likelihood of uttering ill-formed sentences increases as function of the elaborative distance between TS and SS (table 8.4; fig. 8.15).

Point (b) above highlights the importance of always bearing in mind the dynamicity of the system (refer back to 9.1). Status of unit is a matter of entrenchment, which in turn is a function of cognitive salience and frequency (Langacker, 2000, 2008; Dąbrowska, 2004). Status of unit is thus dynamic in nature and consequently, subject to change, as is the way in which a TS is categorised through an SS. Categorisation spans a continuum from extensions which might require more or less effort to recognise a TS as a "distorted" type of a SS, to relationships of elaboration in which the superordinate SS can be more or less fine-grained in its specifications, to total identity of TS and SS.

The dynamicity of the system (and of status of unit in particular) is more clearly inferable from the results of the experimental study. The steady and gradual increase of participants productive with the nonce verb as a function of age can be interpreted as a function of the entrenchment of the transitive constructional schemas *(S)+V<O*. Such a developmental path strongly indicates that full-productivity is a matter of degree and that adultlike syntactic competence emerges in a piecemeal fashion

Such an interpretation of the results, I would argue, is consistent with both previous studies regarding the acquisition of English (fig. 3.9) and the developmental patterns emerging from studies regarding other areas of cognition, such as Theory of Mind and Object Permanence Concept.

If one takes a broader look at developmental evidence, it is possible to note that children initially appear to have a weak representation of various aspects of cognition (false beliefs, object permanence, the transitive construction), which they can exploit in certain "passive tasks" (looking and pointing), but not in other, "more active" ones (telling where Mark will look for his sweet, reaching behaviour, linguistic production). As children gain more and more experience, they become increasingly able to exploit their knowledge in a wider range of experimental tasks.

Munakata et al.'s (1997) simulation model provides evidence that different tasks are likely to require different degrees of representational strength in order to be carried out successfully and that knowledge can indeed be characterised in terms of graded representations. Crucially, their simulation model could form some weak representation of objects from very early on in the learning process. This suggests that positing specific innate bases for children's knowledge is not necessary to account for early competence. Instead, knowledge can be described as a cognitive process that emerges gradually, growing from weak to strong as a function of experience. Thus, results pertaining to the acquisition of the transitive construction can be accounted for in terms of a representation of it that starts out as weak and gradually strengthens, becoming more and more entrenched as children experience more and more language.

16.2.1. TOWARDS A MORE SCHEMATIC REPRESENTATION

A qualitative analysis of Roberto's target sentences and how they are putatively assembled provides an opportunity to speculate on how children may gradually acquire (hence add to their inventories) more schematic units.

For instance, fifteen constructional fails could be accounted for by allowing sentences to be tracked back to three-slot schemas (such as fig. 16.8). The method did not allow three-slot schemas because such units are arguably a bit too schematic to be considered lexically-specific narrow generalisations. Yet, they clearly are not fully-schematic units, either. Nonetheless, if one assumes that the units on which children rely grow in abstraction gradually (as shown by

Dąbrowska, 2000), such units arguably represent one of the most likely steps from a lexically-based organisation of the linguistic system to a more schematic one.

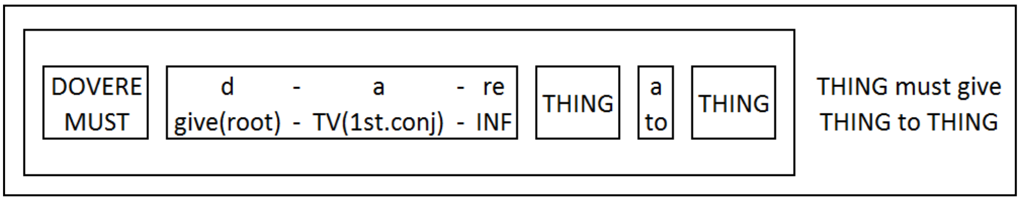


Figure 16.8: A three-slot schema that could account for one of the constructional fails.

At the same time as children develop their lexically-specific units in abstraction (more slots), they may also start to develop fully-schematic units by analogising across lexically-bound schemas. Initially, such fully-schematic units are likely to be weakly entrenched and in these first stages, their retrieval may still depend on the retrieval of lexically-specific units that (at least partially) instantiate them.

The reader may now refer back to the main corpus sentence reported in (52), and fig. 8.4-8.7 (chapter 8.1). The scene experienced by Roberto is one in which E. (X) makes a little dog (Y) fall from a little tractor (Z). This might trigger the categorisation of the event with the unit in fig. 8.6a (because the addressee acts upon something) and fig. 8.6b (because falling is involved). Memory is content addressable and so is the retrieving of linguistic units (see Dąbrowska, 2004). Since the units represented in fig. 8.6a-b partly instantiate the Caused-Motion-Construction (cmc), they provide a pathway via which the child can access the more abstract, less entrenched unit (namely the cmc), which appropriately categorises the whole scene (*X causes Y to move LOCATION_Z*). Once he has done that, the sentence is assembled using the units in fig. 8.6c-f (see fig. 8.7). Thus, the cmc may be retrieved (or accessed) via more specific (partial) instantiations of it.

Clearly, these observations remain nothing but speculations in absence of more carefully designed methods that enquire into how children move from a lexically-specific organisation to a more schematic one. Indeed, this is the research question on which UB researchers must concentrate the most at this point in the research agenda (but see Goldberg, 1999, 2006; Childers & Tomasello, 2001; Dodson & Tomasello, 1998).

16.2.2. THE ROLE OF VOCABULARY, AGE AND COGNITIVE DEVELOPMENT

The analysis of Roberto's language appears to indicate that he is gradually developing more abstract generalisations. Such a claim can be made on the basis of the following observations:

a) Both cases of morph-syntactic overgeneralisations appear to indicate that he is gradually abstracting the Caused-Motion-Construction (chapter 8.1).

b) He acquired the past participle schema *PROCESS-TV-t-GENDER.NUMBER* "PROCESS-TV-ptcp-GENDER.NUMBER".

c) Some of his target sentences could be accounted for by three-slot schemas. Hence, the putative units on which he relies are growing in abstraction.

d) Five Hard_Constructional_Fails could be accounted for by positing that he inserted a component unit within another component unit (see 8.4.6). Lieven et al. (2003) observe that relying on such an operation implies a more mature grasp of constituency (refer back to 9.1.3).

e) He seems to have developed a network of constructions at the nominal level that map onto at least number information (fig. 16.6).

As previously noted, Roberto's vocabulary size is impressively advanced for a twenty-six-month-old. In fact, it would be advanced for a thirty-month-old. It is intriguing that what appears to be the beginning of the path towards a more schematic language is attested in a child with such a large vocabulary.

It is tempting to take such converging outcomes as evidence that the development of a more schematic language is a function of vocabulary growth. Even more so, if one considers that the correlation analyses of the results of the experimental study indicate that vocabulary size is a better predictor of syntactic productivity with the nonce verb than age (table 14.6b; fig. 14.10). The ability to be syntactically productive with the nonce verb can be interpreted as the ability to rely on and retrieve more schematic units. Hence, the relationship between vocabulary and syntactic productivity is expected.

As discussed in the previous chapters, UBAs to linguistic knowledge (e.g. CG) posits that linguistic competence is about mastery of a highly interconnected network of constructions, which posit themselves on a continuum of specificity, schematicity and complexity. Thus, there is not a clear-cut distinction between lexicon and grammar. Learning words' meanings, in particular those of relational words (e.g. *tall, nice*), is inseparable from learning their distributional properties: such knowledge amounts to knowledge of lexically-specific patterns (e.g. *tall_*, *nice_*). Thus, children with a larger vocabulary are more likely to have acquired more lexically-specific patterns; hence, their inventories of lexically-specific constructions are supposedly bigger than those of children with a smaller vocabulary.

According to UB scholars (e.g. Tomasello, 2003), children develop more schematic constructions (*QUALITY THING*) by drawing generalisations out of their more specific units (*tall_*, *nice_*, *terrible_*). Children with a larger vocabulary (and hence a larger inventory of constructions) are more likely to have had more chances to draw generalisations out of their lexically-specific units. Consequently, they are more likely to have developed more schematic units through the necessity of having to store a larger amount of learnt specific patterns (see Goldberg, 1999 and refer back to the discussion in 3.9.2).

The significant relationship between vocabulary and productivity with the nonce verb is therefore consistent with such UB hypotheses on how children move towards a more schematic (adultlike) linguistic knowledge.

However, such conclusions would need further confirmation from other replication studies, especially since, at least in this study, age was a better predictor of syntactic productivity with the **familiar** verb than vocabulary (table 14.6b; fig. 14.10). Indeed, if the results of this study are compared with results regarding English-speaking children, there is a consistent developmental pattern that emerges cross-linguistically. Such a pattern seems to indicate that productivity with nonce material increases as a function of age. Fig. 16.9 compares the proportion of English-speaking children who use nonce verbs in grammatical transitive utterances (from Tomasello, 2006b) with the developmental pattern that emerges from this study.

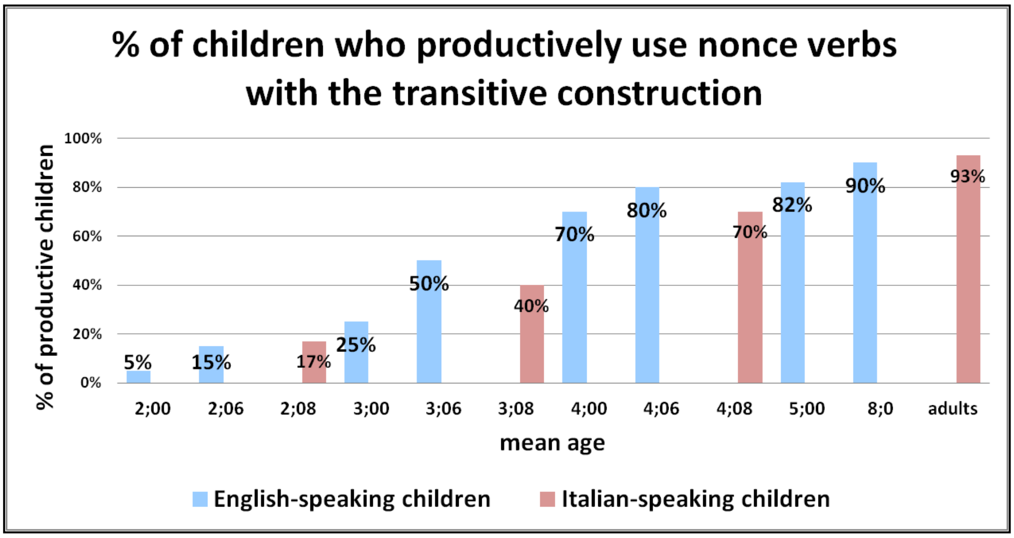


Figure 16.9: proportion of English-speaking children (blue; adapted from Tomasello, 2006b, fig. 6.1, p. 266) and Italian-speaking children (red; refer to fig. 13.3) who produce grammatical transitive utterances with nonce verbs.

Figure 16.9 shows that 15% to 25% of English-speaking children aged 2;06-3;00 and 17% of two-year-olds in this study (M=2;08) are productive with nonce verbs. Performance gets better during the fourth year of life, when about half of the children (40% in this study, 50% of English-speaking children) produce grammatical transitive utterances. One year later, most Italian-speaking (70%) and most English-speaking (70%-80%) children are syntactically productive. This cross-linguistic consistency suggests that age may indeed be a factor in the development of syntactic productivity.

However, it is worth pointing out that the results of this study showed that age and vocabulary are positively related (τ = .53, *p*(two-tailed)<.00001; fig. 14.4). The older children get, the more words (and their associated syntactic patterns) they learn. Thus, it may be the case that studies on English-speaking children have masked the effects of vocabulary size on syntactic productivity as these potential vocabulary effects have not been taken into consideration by many of these previous studies (e.g. Akhtar & Tomasello, 1997; Akhtar, 1999; Abbot-Smith et al., 2001; Lewis, 2009). Thus, since data on children's vocabulary are not available for those studies on English-speaking children, it is not possible to establish which variable (age or vocabulary) is a better predictor of syntactic productivity in those studies.

Nevertheless, the fact that the steps in productivity appear to be similar cross-linguistically calls for further research into the roles of age and vocabulary in the development of syntactic competence.

Children apply, and rely on, their general cognitive abilities in order to acquire and use their ambient language. Hence, it would not be surprising if the acquisition of a pattern as abstract as *AGENT-PROCESS-PATIENT* turned out to be dependent on the development of certain cognitive abilities that, in this case, fully develop around the age of four. A case in point is represented by D'Amico and Devescovi's (1993) and Devescovi et al.'s (1998) studies (refer back to 8.4.5.1). Italian-speaking children younger than five/six years of age do not appear to be capable of disambiguating agent-patient roles on the basis of subject-verb agreement because they have yet to develop the necessary global forms of control in other non-linguistic areas (such as problem-solving and manual-visual coordination). Hence, they are not able to (fully) attend to a global cue like agreement.

As Tomasello (2006b) points out, the study of LA has often been undertaken without the necessary parallel investigation of children's cognitive and social development. In a sense, I would argue, this lack of a more holistic approach is a by-product of the fact that the *nature-vs-nurture* debate has pretty much set the agenda of LA studies, with little or no regard for developmental psychology studies. For many years an innate FL has been the granted, default assumption, and it was up to the "non-believers" to falsify such a claim (see Crain & Petroski, 2001). As a result, UB researchers have (successfully) focussed on showing that children do not have a fully-fledged linguistic knowledge and that there is no (convincing) evidence that they have implicit knowledge of abstract categories such as VP, NP, SUBJECT and the like. A consequence of such a direction taken by LA studies is that little has been done to disentangle how children develop more schematic constructions and (most) linguists have neglected how children's general cognitive development could be used to gain an insight into linguistic ontogeny during early childhood. Nevertheless, if UB researchers aim to unravel how linguistic competence evolves throughout childhood without invoking innate grammatical knowledge, approaches to LA must incorporate both linguistic models and models of cognitive development (Tomasello, 2006b). It indeed appears to be indispensable that LA studies start enquiring into how the development of non-linguistic (cognitive) abilities impact, determine and possibly also constrain children's grammatical development.

16.3. SUMMARY

Some more regular aspects of the morphological system (nominal inflections, past participles) appear to be acquired earlier (during the third year of life) than abstract syntactic patterns such as the transitive construction (acquired during the fifth year of life). Such a developmental sequence seems to hold cross-linguistically and is potentially rooted in the different nature of morphological and syntactic schemas and their instantiations. As for the Italian *passato prossimo* schema, it could be speculated that it presents a recurring concrete element (the morpheme -*t*-) that constantly maps onto a fixed and specific meaning (i.e. past). This creates some kind of semantic and perceptual similarity across the schema's instantiations that is likely to help children's generalisations (fig. 16.2). Such an anchoring of perceptual similarity is hardly found across the instantiations of the transitive construction. Furthermore, the *passato prossimo* schema represents more of a local cue than the transitive schema (hence it is predicted to be acquired earlier on in development) and is also likely to impose a lighter burden on children's working memory (which makes it cognitively less demanding).

Another result that holds cross-linguistically is that vocabulary, rather than age, predicts children's morphological development.

Nevertheless, interesting cross-linguistic differences emerge as to the pace of morphological acquisition: Italian-speaking two-year-olds appear to be more productive with regular past participles than their English-speaking peers. This more advanced morphological behaviour is likely to be the by-product of language specific factors. Firstly, the Italian morphological system is highly regular and frequent, which is a combination that facilitates schema abstraction. Secondly, attending to morphological cues is far more important in Italian than in English. Consequently, Italian-speaking children learn to attend to the communicative functions that morphology has in their ambient language earlier than English-speaking pre-schoolers.

Development of the transitive construction is a gradual and piecemeal process. As children get older, they become increasingly more productive with the nonce verb and the gap in productivity between familiar verb and nonce verb gradually decreases, up to a point (adulthood) in which it vanishes. Such results can be interpreted as evidence of the gradual entrenchment of the transitive schema. Only four-year-olds may be said to have developed a fully-schematic and adultlike competence of it. Conversely, two and three-year-olds' syntactic competence is still better characterised as lexically-specific. This is to some extent confirmed by the longitudinal study: 87% of the syntactic patterns used by Roberto were classified as lexically-specific schemas that could have been inferred from the concrete language he had previously experienced.

Contrasting results are found when the roles of vocabulary and age in syntactic productivity are investigated. Consistent with UB predictions, vocabulary is a better predictor than age with the nonce verb in this study. This suggests that the development of fully-schematic units is a function of linguistic experience. Yet, age seems to be a better predictor than vocabulary when it comes to the familiar verb. Such a result suggests that age may have a bigger role in syntactic development than hypothesised in chapter 12. Indeed, the developmental path that emerges holds cross-linguistically: the transitive construction is acquired by the age of four years (fig. 16.9). Such converging results call for further thorough research into the roles played by age and vocabulary in the development of adultlike syntactic competence. Importantly, it is necessary that future investigations attempt to unravel the extent to which developing such abstract constructions is dependent on more general cognitive abilities and what these abilities may be.

In the meantime, a qualitative analysis of the naturalistic data appears to indicate that the hypotheses UB researchers put forward as to how schema abstraction takes place are on the right track. As children develop their lexically-specific constructions in abstraction (more and more slots), they also start generalising across those units. Abstract templates are inferred by drawing analogies on the basis of formal and functional similarities across more specific patterns. Such analogies might also be facilitated by the fact that light verbs, whose semantics is very general, dominate the token instances of specific Argument Structure

Constructions (Goldberg, 1995, 1999, 2006). Consequently, a specific verb (e.g. *put*) would work as the prototype of a specific constructional pattern (Caused-Motion-Construction), so that the latter inherits its general semantics from the semantics of its prototypical verb-specific instantiation (*X-put-Y-Z*). This in turn boosts schema abstraction: by categorising *I kicked the ball out of the garden* as an extension vis-à-vis *X-put-Y-Z*, a superordinate structure *X-PROCESS-Y-Z* would start entrenching.

Page intentionally left blank

17.

CONCLUSION

The current research has aimed at investigating the extent to which a UBA to LA can be said to have cross-linguistic validity. Such a broad research question has been narrowed down to an investigation into the acquisition of Italian, a highly inflected language whose flexible word order is determined by discourse pragmatics. In doing so, this study has posited itself within a line of research that, since the mid-2000s, has been investigating whether and how well previous results pertaining to English-speaking children generalise to the acquisition of other languages (refer back to 4.1).

17.1. RESEARCH QUESTIONS

In order to shed light on whether the theoretical framework adopted here could account for the acquisition of Italian, two main research questions have been posed:

a) *Can Italian-speaking children's early language be accounted for in terms of lexically-specific units acquired from the concrete language that children themselves have previously experienced?*

b) *To what extent can Italian-speaking children be said to rely on (have mastered) fully-schematic constructions/patterns?*

Research question (a) has been investigated by means of a longitudinal study (Part II) that adopted the traceback method (Lieven et al., 2003, 2009; Dąbrowska & Lieven, 2005) in order to enquire into the spontaneous production of an Italian-speaking two-year-old. The results indicate that the overwhelming (82%) majority of sentences can be accounted for using the traceback method. About 90% of the syntactic patterns Roberto used were classified as lexically-specific units attested twice or more in his own previous linguistic experience (i.e. in the main corpus). Furthermore, the analysis (chapter 8) has shown that virtually all sentences (98%) can find a principled explanation in terms of functional and narrow-scope generalisations from the concrete language he encountered. It is therefore possible to conclude that a UBA can indeed account for the spontaneous production of a two-year-old Italian-speaking child.

An experimental study that tapped into 2;02 to 5;00 year-old Italian-speaking children's morphological (*passato prossimo* construction) and syntactic (transitive construction) productivity with a nonce verb and a familiar verb enquired into the second (b) research question. Experimental evidence points towards two main conclusions. Firstly, all age-groups (two, three and four-year-olds) could be said to be morphologically productive, as each age-group performed similarly across verb_familiarity conditions (at least according to the statistical analysis). Secondly, the same cannot be claimed for their syntactic abilities.

As for morphology, the results appear to indicate that children have developed at least some entrenched schematic morphological knowledge. Nevertheless, two-year-olds were significantly (p<.05) less productive than adults. This suggests that, even though they appeared to be able to bring some schematic knowledge to the experimental task, such knowledge is still not fully adultlike. It is only one year later that children show an adultlike morphological productivity.

Results pertaining to syntactic productivity strongly indicate that the youngest children's syntactic knowledge is better describable as lexically-bound, rather than fully-fledged. The youngest children performed in an adultlike manner with the familiar verb but not with the nonce verb. They also performed significantly better in the former than in the latter verb_familiarity condition. This was not the case for four-year-olds, whose performance was similar across verb_familiarity conditions and adultlike with both verbs. Hence, there seems to be (little or) no evidence of reliance on fully-schematic patterns in the syntactic productivity of two-year-olds and three-year-olds. Conversely, four-year-olds may be said to have acquired a fully-schematic representation of the transitive construction, of which they appear to have an adultlike representation. The results also clearly indicate that the development (entrenchment) of schematic templates is a matter of degree. Children develop a fully-schematic representation of the transitive schema only gradually, in a piecemeal fashion. As they gain more and more linguistic experience and their cognitive abilities mature, their schematic units become more and more entrenched. The result is an increasing proportion of productive participants as a function of age. Such an outcome is consistent with results pertaining to English-speaking children (Akhtar & Tomasello, 1997; Akhtar, 1999), as the developmental trends that emerge in both languages are strikingly similar (fig. 16.9).

A further issue investigated by the experimental study was the roles vocabulary and age play in the development of grammatical competence. As previously discussed, such a research question has important theoretical implications. To summarise what has been argued throughout this study, a UBA predicts that vocabulary and grammar develop hand in hand, as lexicon and grammar represent a continuum whose borders overlap quite substantially.

Vocabulary is indeed a better predictor of morphological productivity than age. Importantly, such results seem to indicate that vocabulary does not only play a crucial role in the age range (0;10 to 2;08) previously investigated by Bates and colleagues (Marchman & Bates, 1994; Caselli et al., 1999; Bates & Goodman, 2001; Devescovi et al., 2005), but it indeed continues to be crucial to grammatical development throughout the pre-school years (2;02 to 5;00), even in the acquisition of Italian. As for the relationship between vocabulary and syntactic development, results are not consistent; vocabulary is the best predictor of productivity with the nonce verb, whereas age is the best predictor of productivity with the familiar verb. Consequently, results do not allow certain conclusions and call for further research.

17.2. FUTURE DIRECTIONS

Although the experimental investigation suggested that, when group productivity is analysed, the youngest children (two-year-olds) were morphologically as productive with the familiar verb as with the nonce verb, such an outcome deserves further replication studies. Firstly, it may be possible that, were more participants recruited, the difference in productivity between the two verbs (-32%) would reach significance. Secondly, the extent to which results would generalise to other Conjugation I inflections and/or to other inflectional classes (Conjugation II and III) is still unclear. Further research should therefore address this issue by focussing on productivity with both singular and plural inflections of both Conjugation I and III regular nonce verbs.

Another issue that deserves further investigation is the roles of age and vocabulary in the development of a schematic syntactic representation. Future research will have to better control for the interaction between children's ToM development and the pragmatics behind argument omission in early Italian. A further observation is that studies that adopt the nonce verb technique must start measuring participants' vocabulary, as none of the studies reviewed here appear to have done so. Clearly, this is essential if a more precise insight into the portion of unique variance for which each variable accounts is to be gained.

Most importantly, UB researchers need to now focus their efforts on developing research methods which investigate how exactly children develop a more adultlike representation of various morpho-syntactic constructions out of more specific units. Importantly, approaches to LA must more seriously start investigating which non-linguistic cognitive abilities are putative prerequisites to schematise those specific morpho-syntactic constructions.

17.3. THE LARGER PICTURE

In the introductory chapter (0), it was pointed out that studying LA can help us to gain an insight into what language, a complex mental and cognitive system, is. Specifically, studying LA can shed light on language ontogeny, productivity and learnability. Clearly the three issues are related and overlap.

Without going into too much detail, some broad observations can be made on the basis of the results yielded by this research. The naturalistic study suggests that children start out by drawing narrow, lexically-specific and functional generalisations on the basis of which they infer the putative symbolic units of their ambient language (learnability). Importantly, this implies that the input children receive is rich enough to allow children to draw these generalisations from the concrete language they experience. As a consequence, they can exploit such generalisations to produce and understand (through relationships of categorisation, composition and symbolisation) many novel utterances (productivity). Both the qualitative analysis of the naturalistic study and the results of the experimental investigation appear to suggest that children slowly move from a lexically-specific knowledge to a more schematic and adultlike one (ontogeny).

Overall, it appears safe to conclude that a UBA to LA has cross-linguistic validity, as it can account for the acquisition of Italian in both naturalistic and experimental settings.

Page intentionally left blak

18.

REFERENCES

Abbot-Smith, K. & Behrens, H. (2006) How Known Constructions Influence the Acquisition of Other constructions: The German Passive and Future Constructions. *Cognitive Science*, 30, 995-1026.

Abbot-Smith K, Lieven, E.V.M. & Tomasello M. (2001) What preschool children do and do not with ungrammatical word orders. *Cognitive Development*, 16(2), 679-692.

Abbot-Smith, K. & Serratrice, L. (2013) Word order, referential expression, and case cues to the acquisition of transitive sentences in Italian. *Journal of Child Language*, 1, 1-31.

Aguado-Orea, J.J. (2004) *The acquisition of morpho-syntax in Spanish: implications for current theories of development*. Unpublished PhD thesis, University of Nottingham.

Aitchison, J. (2011) *The Articulate Mammal (6th ed.)*. London and New York: Routledge. Epub edition.

Airenti, G. (2015) Theory of mind: a new perspective on the puzzle of belief ascription. *Frontiers in Psychology*, 6, 1184. http://doi.org/10.3389/fpsyg.2015.01184.

Akhtar, N. (1999) Acquiring basic word order: evidence for data-driven learning of syntactic structure. *Journal of Child Language*, 26, 339-356.

Akhtar, N. & Tomasello M. (1997) Young Children's Productivity with Word Order and Verb Morphology. *Developmental Psychology*, 33, 952-965.

Allen, S. (1997) A discourse-pragmatic explanation for the subject-object asymmetry in early null arguments: The principle of informativeness revisited. In A. Sorace, C. Heycock & R. Shillcock (eds) *Proceedings of the GALA '97 Conference on Language Acquisition* (pp. 10–15). Edinburgh: Human Communication Research Centre.

Allen, S. (2000) A discourse-pragmatic explanation for argument representation in child Inuktitut. *Linguistics*, 38, 483-521.

Ambridge, B. & Lieven, E.V.M. (2011) *Child Language Acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press.

Ambridge, B. & Rowland, C. (2009) Predicting children's errors with negative questions: Testing a schema-combination account. *Cognitive Linguistics*, 20(2), 225-266.

Ambridge, B., Rowland, C.F. & Pine, J.M. (2008) Is Structure Dependence an Innate Constraint? New Experimental Evidence From Children's Complex-Question Production. *Cognitive Science*, 32, 222-255.

Anjomshoa, L. & Sadighi, F. (2015) Minimalism, Economy, Simplicity, and Children Language Acquisition. *Social and Behavioural Sciences*, 192, 142–149.

Antelmi, D. (1997) *La prima grammatica dell'Italiano*. Bologna: Il Mulino.

Antinucci, F. & Cinque, G. (1977) Sull'ordine delle parole in italiano: l'emarginazione. *Studi di Grammatica Italiana*, 6, 121-146.

Baayen, R.H. (2008) *Analysing Linguistic Data. A Practical Introduction to Statistics using R*. New York: Cambridge University Press.

Bannard, C., Lieven, E.V.M. & Tomasello,M. (2009) Modeling children's early grammatical knowledge, *Proceedings of the National Academy of Sciences, 106* (41), 17284-17289.

Barbieri, G. (1971) *Le strutture della nostra lingua*. Firenze: La Nuova Italia Editrice.

Bartsch K. & Wellman, H.M. (1995) *Children talk about the mind*. New York, NY: Oxford University Press.

Bates, E. (1976) *Language and Context: The Acquisition of Pragmatics*. New York: Academic Press.

Bates, E., Bretherton, I. & Snyder, L. (1988) *From First Words to Grammar: Individual Differences and Dissociable Mechanisms*. New York: Cambridge University Press

Bates, E. & Goodman, J.C. (2001) On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition. In M. Tomasello and E. Bates (eds) *Language Development. The Essential Readings* (pp. 134-162). Malden, Mass. and Oxford, UK: Blackwell Publishers.

Bates, E. & MacWhinney, B. (1987) Competition, Variation, and Language Learning. In B. MacWhinney (ed.), *Mechanisms of Language Acquisition* (pp. 157-193). Hillsdale, NJ and London: Lawrence Erlbaum.

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F. & Venza, V. (1984) A Cross-Linguistic Study of the Development of Sentence Interpretation Strategies. *Child Development*, 55(2), 341-354.

Bates, D.M. (2010) *lme4: Mixed-effects modelling with R*. Springer. Retrieved at http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf

Bauer, P. (1996) What do infants recall of their lives? Memory for specific events by 1-to-2-year-olds. *American Psychological Association*, 51, 29-41.

Belletti, A. (2004) Aspect of the Low Ip Area. In L. Rizzi (ed.) *The structure of CP and IP* (pp. 16-51). New York: Oxford University Press.

Benincà, P. (2001) The position of Topic and Focus in the Left Periphery. In G. Cinque & G. Salvi (eds) *Current Studies in Italian Syntax. Essays offered to Lorenzo Renzi* (pp. 39-64). Amsterdam: Elsevier.

Benincà, P. with Salvi, G. & Frison, L. (2001) L'ordine della frase e le costruzioni marcate. In L. Renzi, G. Salvi, & A. Cardinaletti (eds) *Grande grammatica italiana di consultazione (2nd ed.)*, vol. I (pp. 129-239). Bologna: Il Mulino.

Berk, A.R. (2006) An Introduction to Ensemble Methods for Data Analysis. *Sociological Methods Research*, 34(3), 263-295.

Bever (1970a) The Cognitive basis for linguistic structures. In J.R. Hayes (ed.) *Cognition and the Development of Language* (pp. 279-352). New York: Wiley.

Bever (1970b) The Integrated Study of Language. In J. Morton (ed.) *Biological and Social Perspective on Language* (pp. 158-206). London: Logos Press.

Bickerton, D. (2009) Recursion: core of complexity or artefact of analysis? In T. Givón & M. Shibatan (eds) *Syntactic complexity: diachrony, acquisition, neuro-cognition, evolution* (pp. 532-544). Amsterdam: John Benjamins.

Boeckx, C. (2006). *Linguistic Minimalism. Origins, Concepts, Methods and Aims*. Oxford: Oxford University Press.

Booij, G. (2010) Construction Morphology. *Language and Linguistics Compass*, 3(1), 1-13.

Bortolini, U., Tagliavini C. & Zampolli, A. (1971). *Lessico di frequenza della lingua italiana contemporanea*. Milan: Garzanti.

Boyd, J.K. & Goldberg, A. (2012) Young children fail to fully generalize a novel argument structure construction when exposed to the same input as older learners. *Journal of Child Langauge*, 39, 457-481.

Braine, M. (1963) The ontogeny of English phrase structure. *Language*, 39, 1-14.

Braine, M. (1976) Children's first word combinations. *Monographs of the Society for Research in Child Development*, 41 (Serial No. 167), 1-104.

Brooks, P.J. & Tomasello (1999) Young Children Learn to Produce Passives With Nonce Verbs. *Developmental Psychology*, 35(1), 29-44.

Brown & Kane (1988) Preschool children can learn to transfer: Learning to learn and learning from examples. *Cognitive Psychology*, 20, 493-523.

Camaioni, L. (2001) Lo sviluppo del lessico. In L. Camaioni (ed.) *Psicologia dello Sviluppo della Lingua* (pp. 85-115). Il Mulino: Bologna.

Caselli, M.C. & Casadio, P. (1995) *Il primo vocabolario del bambino: guida all'uso del questionario MacArthur per la valutazione della comunicazione e del linguaggio nei primi anni di vita*. Milan: Franco Angeli Editore.

Caselli, M.C., Casadio, P. & Bates, E. (1999). A comparison of the transition from first word to grammar in English and Italian. *Journal of Child Language*, 26, 69-111.

Cardinaletti, A. (2001) A Second Thought on *Emarginaione*: Destressing vs. "Right Dislocation". In G. Cinque and G. Salvi (eds) *Current Studies in Italian Syntax. Essays Offered to Lorenzo Renzi* (pp. 65-135). Amsterdam: Elsevier.

Chan, A., Lieven, E., & Tomasello, M. (2009) Children's understanding of the agent–patient relations in the transitive construction: cross-linguistic comparison between Cantonese, German and English. *Cognitive Linguistics*, 20(2), 267-300.

Childers, J.B. & Tomasello, M. (2001) The Role of Pronouns in Young Children's acquisition of the English Transitive Construction. *Developmental Psychology*, 37(6), 739-748.

Chipere, N. (2001) Native speaker variations in syntactic competence: implications for first language teaching. *Lang. Aware*, 10, 107–124.

Chomsky, N. (1959) A Review of B. F. Skinner's *Verbal Behavior* (1957). *Language,* 25(1), 26-58. In L. A. Jakobovitsn & M. S. Miron (eds) *Readings in the Psychology of Language*, 1967. Retrieved at https://chomsky.info/1967____/.

Chomsky, N. (1965) *Aspect of the Theory of Syntax*. Cambridge, Massachusetts: M.I.T. Press.

Chomsky, N. (1972). *Language and Mind*. New York, NY: Harcourt Brace Jovanovich.

Chomksy, N. (1975) *Reflections on Langauge*. London: Temple Smith.

Chomsky, N. (1980) Rules and representations. *Behavioral and Brain Sciences,* 3, 1–61.

Chomsky, N. (1986) *Knowledge of Language: its Nature, Origins and Use*. New York: Praeger.

Chomsky, N. (1995) *The Minimalist Program*. Cambridge, Massachusetts and London: MIT Press.

Chomsky, N. (2000) *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. (2004) Beyond Explanatory Adequacy. In A. Belletti (ed.) *Structures and Beyond*. *The Cartography of Syntactic Structure, vol. III* (pp. 103-131). Oxford University Press, New York.

Chomsky, N. (2005) Three factors in language design. *Linguistic Inquiry*, 36(1), 1–22.

Chomsky, N. (2007) Approaching UG from Below. In U. Sauerland & H.M. Gartner (eds) *Interfaces + Recursion = Language?: Chomsky's Minimalism and the View from Syntax-Semantics* (pp. 1-29). Berlin and New York: Mouton de Gruyter.

Cianchetti, C. & Sannio Facello, G. (2010) *Test di Valutazione del Linguaggio. Livello Prescolare*. Trento: Edizioni Erickson.

Cipriani, P. Chilosi, A. & Bottari, P. (1993) Inferenza strutturali nell'acquisizione della morfologia libera italiana. In E. Cresti & M. Moneglia, (eds.) *Ricerche sull'acquisizione dell'italiano*. *Giornnata di studio sull'acquisizione del linguaggio in bambini normali e patologici (Firenze, 30 maggio 1991)* (pp. 189-215). Roma: Bulzoni.

Cipriani, P, Chilosi, A., Bottari, P. & Planner, L. (1993) *L'acquisizione della Morfosintassi in Italiano. Fasi e processi*. Padova: Unipress.

Cipriani, P., Pfanner, P., Chilosi, A., Cittadoni, L., Ciuti, A., Maccari, A., Pantano, N., Pfanner, L., Poli, P., Sarno, S., Bottari, P., Cappeli, G., Colombo, C., & Veneziano, E. (1989). *Protocolli diagnostici e terapeutici nello sviluppo e nella patologia del linguaggio* (1/84 Italian Ministry of Health). Pisa: Stella Maris Foundation.

Clark, R. (1974) Performing without competence. *Journal of Child Language*, 4, 341-358.

Cook, V. J. & Newson, M. (2007) *Chomsky's Universal Grammar: An Introduction (3rd ed.)*. Malden, MA: Blackwell.

Cordin, P. (2001) Il clitico «ne». In L. Renzi, G. Salvi & A. Cardinaletti (eds) *Grande grammatica italiana di consultazione (2nd ed.)*, vol. I (pp. 647-710). Bologna: Il Mulino.

Cordin, P. & Calabrese, A. (2001) I pronomi personali. In L. Renzi, G. Salvi & A. Cardinaletti (eds) *Grande grammatica italiana di consultazione (2nd ed.)*, vol. I (pp. 549-606). Bologna: Il Mulino.

Crain S. & Pietroski, P. (2001) Nature, Nurture and Universal Grammar. *Linguistics and Philosophy*, 24(2), 139-186.

Crawley, M. J. (2007). *The R book*. New York: Wiley.

Croft, W. & Cruise, D.A. (2004) *Cognitive Linguistics*. New York: Cambridge University Press.

Dąbrowska, E. (2000) From formula to schema: The acquisition of English questions. *Cognitive Linguistics*, 11, 83-102.

Dąbrowska, E. (2004) *Language, Mind and Brain: Some Psychological and Neurological Constraints on Theories of Grammar*. Edinburgh: Edinburgh University Press.

Dąbrowska, E. (2008) The later development of an early-emerging system: The curious case of the Polish genitive. *Linguistics,* 46, 629–650.

Dąbrowska, E. (2009) Words as constructions. In E. Evans & S. Pourcel (eds) *New directions in cognitive linguistics* (pp. 201-223). Amsterdam: John Benjamins.

Dąbrowska, E. (2012) Different speakers, different grammars: individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, 2(3), 219–253.

Dąbrowska, E. (2014) *Recycling utterances: A speaker's guide to sentence processing.* Cognitive Linguistics, 25(4), 617-653.

Dąbrowska E. (2015) What exactly is Universal Grammar, and has anyone seen it? *Front. Psychol.*, 6:852. doi: 10.3389/fpsyg.2015.00852.

Dąbrowska, E. & Lieven, E.V.M. (2005) Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics,* 16(3), 437-474.

Dąbrowska, E. & Tomasello, M. (2008) Rapid learning of an abstract language-specific category: Polish children's acquisition of the instrumental construction. *Journal of Child Language*, 35, 335-558.

D'Amico, S. & Devescovi, A. (1993) Processi di comprensione dei bambini italiani: l'interpretazione della frase semplice. In E. Cresti & M. Moneglia (eds), *Ricerche sull'acquisizione dell'italiano*. *Giornnata di studio sull'acquisizione del linguaggio in bambini normali e patologici (Firenze, 30 maggio 1991)* (pp. 273-289). Roma: Bulzoni.

Dell'Orletta F, Lenci, A., Montemagni, S. & Pirrelli, V. (2005) Climbing the path to grammatical: A maximum entropy model of subject/object learning. *Second Workshop on Psychocomputational Models of Human Language Acquisition*, pp. 72–81. Retrived at http://www.aclweb.org/anthology/W05-0509.

De Mauro, T., Mancini, F., Vedovelli, M. & Voghera, M. (1993) *Lessico di frequenza dell'italiano parlato*. Milan: Etas Libri.

Devescovi, A., Caselli, M., Marchione, D., Pasqualetti, P., Reilly, J. & Bates, E. (2005) A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language*, 32, 759-786.

Devescovi, A. & D'Amico, S. (2001) Lo sviluppo della morfosintassi. In L. Camaioni (ed.) *Psicologia dello Sviluppo del Linguaggio* (pp. 117-152). Il Mulino: Bologna.

Devescovi, A, D'Amico, S, Smith, S., Mimica, I & Bates, E. (1998) The development of sentence comprehension in Italian and Serbo-Croatian: local versus distributed cues. In F. Hillert (ed.) *Syntax and Semantics, Vol. 31. Sentence Processing: A cross-linguistic Perspective* (pp. 345-377). New York: Academic Press.

Di Eugenio, B. (1990) Centering theory and the Italian pronominal system, *Proceedings of the 13th International Conference on Computational Linguistics (COLING '90)* (pp. 270–275). Helsinki, Finland. Retrieved at http://www.aclweb.org/anthology-new/C/C90/C90-2047.pdf

Dodson, K. & Tomasello, M. (1998) Acquiring the transitive construction in English: the role of animacy and pronouns. *Journal of Child Language*, 25, 555-574.

D'Odorico L., Fasolo M., Cassibba R., & Costantini A. (2011) "Lexical, Morphological, and Syntactic Characteristics of Verbs in the Spontaneous Production of Italian Children," *Child Development Research*, vol. 2011, Article ID 498039, 17 pages. doi:10.1155/2011/498039. Retrieved at http://www.hindawi.com/journals/cdr/2011/498039/cta/

Dressler, W.U., & Thornton, A.M. (1991) Doppie basi e binarismo nella morfologia italiana. *Rivista di Linguistica*, 2, 3–22.

Field, A., Miles, J. & Field, Z. (2012) *Discovering Statistics using* R (epub edition). Los Angeles, London, New Deli, Singapore and Washington: Sage.

Fisher, C. (2002) The role of abstract syntactic knowledge in language acquisition: a reply to Tomasello (2000). *Cognition*, 82, 259-278.

Gentner, D. & Markman, A. (1995) Similarity is like analogy: Structural alignment in comparison. In C. Cacciari (ed.) *Similarity in language, thought and perception* (pp. 111-147). Brussels: BREPOLS.

Gentner, D. & Markman, A. (1997) Structure mapping in analogy and similarity. *American Psychologist*, 52, 45-56.

Gentner, D. & Medina, J. (1998) Similarity and the development of rules. *Cognition*, 65, 263-297.

Gass, M.S. (1987) The resolution of conflicts among competing systems: A bidirectional perspective. *Applied Pscycholinguistics*, 8, 329-350.

Goldberg, A.E. (1995) *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago and London: University of Chicago Press.

Goldberg, A.E. (1999) The emergence of the semantics of argument structure constructions. In B. MacWhinney (ed.) *The emergence of language* (pp. 197-212). Mahwahn, NJ: Laurence Erlbaum Associates.

Goldberg, A.E. (2006) *Constructions at Work. The Nature of Generalization in Language*. New York: Oxford University Press.

Goldberg, A.E, Casenhiser D.M. & Sethuraman, N. (2004) Learning Argument Structure Generalisations. *Cognitive Linguistics*, 15(3), 289-316.

Guasti, M.T. (1993/94) Verb syntax in Italian Child Grammar: Finite and Nonfinite Verbs. *Language Acquisition*, 3(1), 1-40.

Guasti, M.T. (2001) La struttura interna del sintagma aggettivale. In L. Renzi, G. Salvi & A. Cardinaletti (eds) *Grande grammatica italiana di consultazione (2nd ed.)*, vol. II (pp. 321-337). Bologna: Il Mulino.

Guasti, M.T., (2002) *Language Acquisition: The growth of Grammar.* Cambridge, Massachusetts: MIT Press.

Haegeman, L., (1994) *Introduction to Government and Binding*. Oxford: Blackwell.

Harris, P.L. (2006) Social Cognition. In D. Kuhn & R.S. Siegler (eds) *Cognition, perception, and language (6th ed.)*. In W. Damon & R.M. Lerner (eds) *Handbook of child psychology*, vol. 2 (pp. 811-857). Hoboken, NJ: Wiley.

Hauser, M., Chomsky, N., & Fitch, T. (2002) The faculty of language: what is it, who has it, and how did it evolve. *Science* 198, 1569-79.

Hebb, D. O. (1961) *The Organization of Behavior: A Neuropsychological Theory*. New York: John Wiley & Sons. Science Editions.

Hirsh-Pasek, K. & Michnick Golinkoff, R. (1998) Elicited Production. In D. McDaniel, C. McKee & H. Smith (eds) *Methods for Assessing Children's Syntax* (pp. 77-102). Cambridge, Massachusetts and London: MIT Press.

Hornstein, N., Nunes, J. & Grohmann, K.K. (2005) *Understanding Minimalism*. Cambridge: Cambridge University Press

Hothorn T., Hornik, K. & Zeileis, A. (2006) Unbiased recursive partitioning: A conditional inference framework. *Journal of computational and Graphics Statistics 2006*, 15(3), 382-417.

Hothorn, T., Hornik, K. & Zeileis, A. (*cran.r_project*) ctree: Conditional Inference Trees. *Cran.r_project*. Retrieved at http://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf.

Hunt, K. (1965). *Grammatical Structures Written at Tree Grade Level*s. NCTE Research report No. 3. Champaign, IL, USA: NCTE.

Johnson, K. (2008) *Quantitative Methods in Linguistics*. Oxford: Blackwell Publishing.

Jusczyk, P.W. (1997) *The Discovery of spoken Language*. Cambridge, Massachusetts: MIT Press.

Kail, M. & Charvillat, A. (1998) Local and Topological Processing in Sentence Comprehension by French and Spanish Children. *Journal of Child Language*, 15, 637-662.

Kavanaugh R.D. (2006) Pretend play and theory of mind. In L. Balter & C.S. Tamis-LeMonda (eds) *Child psychology: A handbook of contemporary issues (2nd ed.)* (pp. 155-166). New York: Psychology Press.

Kemler Nelson, D. (1974) Concept, word & sentence: Interrelations in acquisition and development. *Psychological Review*, 81(4), 267-285.

Kelmer Nelson, D., Russel, R., Duke, N. & Jones, K. (2000) Two-Year-Olds Will Name Artifacts by Their Functions. *Child Development*, 71(5), 1271-1288.

Kenneth, R. (1983) Comprehension and Production of Word Order in stage I. *Child Development*, 54 (2), 443-449.

Kucera, H & Francis, N. (1967). *Computational analysis of present-day English*. Providence (RI): Brown University Press.

Langacker, R.W. (1987) *Foundation of Cognitive Grammar. Vol. I: Theoretical Prerequisites*. Stanford: Stanford University Press.

Langacker, R.W. (1991) *Foundation of Cognitive Grammar. Vol. II: Descriptive Application*. Stanford: Stanford University Press.

Langacker, R.W. (2000) A Dynamic Usage-Based Conception of Language. In M. Barlow & S. Kemmer (eds) *Usage-based Models of Language* (pp. 1-61). Cambridge: Cambridge University Press.

Langacker, R.W. (2008) *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press.

Langacker, R.W (2010) How to agree not to disagree. The emergence of structure from usage. In K. Boye & E. Engberg-Pedersen (eds) *Language Use and Language structure* (pp. 107-144). Berlin and New York: Mounton de Gruyter.

Larson-Hall, J. (2010a) *A Guide to Doing Statistics in Second Language Research Using R*. Routledge, Taylor & Francis. Retrieved at http://cw.routledge.com/textbooks/9780805861853/guide-to-r.asp

Larson-Hall, J. (2010b) *A Guide to Doing Statistics in Second Language Research Using SPSS*. New York and London: Routledge, Taylors & Francis.

Lehmann, C. (1982) Directions for interlinear morphemic translations. *Folia Linguistica* 16, 199-224.

Leonard, L.B., Caselli, M.C. & Devescovi, A. (2002) Italian children's use of verb morphology during the preschool years. *First Language*, 22, 387-304.

Lepschy, L. & Lepschy, G. (1977) *La lingua italiana.* Milano: Bompianti.

Lewis, L. (2009) Novel verb productivity in young English and Spanish speakers. *Journal of Psychological Arts and Sciences, 1,* 1-5.

Lieven E.V.M. & Brandt S. (2011) The constructivist approach. *Infancia y aprendizaje*, 34(3), 281-296.

Lieven, E.V.M. & Behrens, H. (2012) Dense Sampling. In E. Hoff (ed) *Research Methods in Child Language. A Practical Guide* (pp. 226–239). Oxford: Wiley-Blackwell.

Lieven, E.V.M., Behrens, H., Speares, J. & Tomasello M., (2003) Early syntactic creativity: a usage-based approach. *Journal of Child Language*, 30(2), 333-367.

Lieven, E.V.M., Pine, J.M. & Baldwin, G. (1997) Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187-219.

Lieven, E.V.M, Salomo, D. & Tomasello, M. (2009) Two-year old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics,* 20(3), 481-507.

Lieven, E.V.M. & Tomasello, M. (2008) Children's first language acquisition from a usage-based perspective. In P. Robinson & N.C. Ellis (eds) *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 168-196). New York: Routledge.

Longa, V. & Lorenzo, G. (2008) What about a (really) minimalist theory of language acquisition. *Linguistics*, 46(3), 541-570.

Lorusso, P., Caprini, C. & Guasti, M.T. (2005) Overt Subject Distribution in Early Italian Children. *Online Supplement of the Boston University Conference on Language Development*, vol. 29. Eds: A. Brugos, M.R. Clark-Cotton & S. Ha. Retrieved at http://www.bu.edu/bucld/proceedings/supplement/vol29/

Lust, B. (2006), *Child Language: acquisition and growth*. Cambridge: Cambridge University Press.

Marchman, V. & Bates, E. (1994) Continuity in lexical and morphological development: a test of the critical mass hypothesis. *Journal of Child Language*, 21, 339-366.

MacWhinney, B. (2000) *The CHILDES Project: Tools for Analyzing Talk (3rd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Matracki, I.P. (2010) L'accordo e l'ordine delle parole in italiano. *Studia Romanica et Anglica Zagrabiensia*, 54, 51-82.

Matthews, D., Lieven, E.V.M., Theakston, A. & Tomasello, M. (2007) French children's use and correction of weird word order: A constructivist account. *Journal of Child Language*, 34, 381-409.

McClure, K., Pine, J.M. & Lieven, E.V.M. (2006) Investigating the abstractness of children's early knowledge of argument structure. *Journal of Child Language*, 33, 693-720.

Munakata Y., McClelland, J.L., Johnson, M.H. & Siegler, R.S. (1997) Rethinking Infant Knowledge: Towards an Adaptive Process Account of Successes and Failures in Object Permanent Tasks. *Psychological Review*, 104(4), 686-713.

Nagakawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Ecology and Evolution*, 4(2), 133-142.

Nespor, M. (2001) Il Sintagma Aggettivale. In L. Renzi, G. Salvi & A. Cardinaletti (eds.) *Grande grammatica italiana di consultazione (2nd ed.)*, vol. I (pp. 439-455). Bologna: Il Mulino.

Nippold, M.A., Hesketh, L.J., Duthie, J.K. & Mnasfield, T.C. (2005) Conversational Versus Expository Discourse: A Study of Syntactic Development in Children, Adolescents, and Adults. *Journal of Speech, Language and Hearing Research*, 48, 1048-1064.

Noble, C., Rowland, C. & Pine, J.M. (2011) Comprehension of Arguments Structure and Semantic Roles: Evidence from English-Learning Children and the Forced-Choice Pointing Paradigm. *Cognitive Science*, 35, 963-982.

O'Grady (2012) Three Factors in the Design and Acquisition of Language. *Cogn Scie*, 3, 493-499. Doi: 10.1002/wcs.1188.

Orsolini, M., & Marslen-Wilson, W. (1997) Universals in morphological representation: Evidence from Italian. *Language and Cognitive Processes*, 12, 1–47.

Orsolini, M., Fanari, R., & Bowles, H. (1998) Acquiring regular inflection in a Language with Verb Classes. *Language and Cognitive Processes*, 13(4), 425-464.

Piaget, J. (1952) *The Origins of Intelligence in Children*. New York: Norton (original work published in 1935).

Peterson, C.C., Wellman, H.M. & Slaughter, V. (2012) The Mind Behind the Message: Advancing Theory-of-Mind Scales for Typically Developing Children, and Those With Deafness, Autism, or Asperger Syndrome. *Child Development,* 83, 469–485. doi:10.1111/j.1467-8624.2011.01728.x

Pine, J.M. & Lieven, E.V.M. (1993) Reanalysing rote-learned phrases: individual differences in the transition to multi-word speech. *Journal of Child Language*, 20, 551-571.

Pine, J.M. & Lieven, E.V.M. (1997) Slot and Frames Patterns and the Development of the Determiner Category. *Applied Psycholinguistics*, 18, 123-138.

Pizzuto, E. & Caselli, C. (1992) The acquisition of Italian morphology: implications for models of language development. *Journal of Child Langauge*, 19, 491-557.

Pizzuto, E. & Caselli, C. (1993) L'Acquisizione della morfologia flessiva nel linguaggio spontaneo: evidenza per modelli innatisti o cognitivisti? In E. Cresti & M. Moneglia (eds), *Ricerche sull'acquisizione dell'italiano. Giornata di studio sull'acquisizione del linguaggio in bambini normali e patologici (Firenze, 30 maggio 1991)* (pp. 165-187). Roma: Bulzoni.

Pizzuto, E. & Caselli, C. (1994) The acquisition of Italian verb morphology in a cross-linguistic perspective. In Y. Levy (ed.) *Other Children Other Languages. Issues in the Theory of Language acquisition* (pp. 137-188). Hillsdale, N.J.: Erbaum.

Radden, G. & Driven, R. (2007) *English Cognitive Grammar*. Amsterdam and Philadelphia: John Benjamins.

Radford, A. (1990) *Syntactic theory and the acquisition of English Syntax*. Oxford: Basil Blackwell.

Radford, A. (2009) *An Introduction to English Sentence Structure*. Cambridge: Cambridge University Pres.

Rizzi, L. (2005) On the Grammatical Basis of Language Development: A Case Study. In G. Cinque G. & R.S. Kayne (eds) *The Oxford Handbook of Comparative Syntax* (pp. 70-109). Oxford: Oxford University Press.

Rubino, R. & Pine, J. (1998) Subject-verb agreement in Brazilian Portuguese: What low error rates hide. *Journal of Child Language*, 25, 35-60.

Russo, J.L. (1954) *Practical Italian Grammar*. London and Edinburgh: George & G. Harrap & co. LTD.

Salvi, G. (2001a) La frase semplice. In L. Renzi, G. Salvi & A. Cardinaletti (eds.) *Grande grammatica italiana di consultazione (2nd ed.)*, vol. I (pp. 37-239). Bologna: Il Mulino.

Salvi. G. (2001b) L'accordo. In L. Renzi, G. Salvi & A. Cardinaletti (eds.) *Grande grammatica italiana di consultazione (2nd ed.)*, vol. II (pp. 227-244). Bologna: Il Mulino.

Say, T. & Clahsen, H. (2002) Words, Rules and Stems in the Italian Mental Lexicon. In S. Nooteboom, F. Weerman & F. Wijnen (eds.) *Storage and Computation in the Language Faculty. Studies in Theoretical Psycholinguistics, vol. 30* (pp. 93-129). Retrieved May, 2013, from http://www.unipotsdam.de/fileadmin/projects/prim/papers/italo.html.

Serratrice, L. (2005) The role of discourse pragmatics in the acquisition of subjects in Italian. *Applied psycholinguistics*, 26, 437-462.

Serratrice, L., Sorace, A. & Paoli, S. (2004) 'Crosslinguistic influence at the syntax-pragmatics interface: Subjects and objects in English-Italian bilingual and monolingual acquisition'. *Bilingualism and Cognition*, 7, 183-205.

Scalisse, S. (2001) La formazione delle parole. In L. Renzi, G. Salvi & A. Cardinaletti (eds), (2001) *Grande grammatica italiana di consultazione*, vol. III (pp. 473-516). Bologna: Il Mulino.

Schwartz, B., & Sprouse, R. (2013). Generative approaches and the poverty of the stimulus. In J. Herschensohn & M. Young-Scholten (eds) *The Cambridge Handbook of Second Language Acquisition* (pp. 137-158). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139051729.010

Slaughter, V. (2015) Theory of Mind in Infants and Young Children: A Review. *Australian Psychologist*, 50, 169–172. doi:10.1111/ap.12080

Slobin, D. (1982) Universal and particular in the acquisition of language. In E. Wanner & L. Gleitman (eds) *Language acquisition: The state of the art* (pp. 128-170). Cambridge: Cambridge University Press.

Slobin, D. & Bever, T.G (1982) Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, 12, 229-265.

Street, J. (2017) This is the native speaker that the non-native speaker outperformed: Individual, education-related differences in the processing and interpretation of Object Relative Clauses by native and non-native speakers of English. *Language Sciences*, 59, 192–203.

Street, J. & Dąbrowska, E. (2010) More individual differences in language attainment: how much do adult native speakers of English know about passives and quantifiers? *Lingua*, 120, 2080–2094.

Street, J. & Dąbrowska, E. (2014) Lexically specific knowledge and individual differences in adult native speakers' processing of the English passive. *Appl. Psycholinguist*., 35, 97–118.

Strobl, C., Mallesy, J. & Tutz, G. (2009) An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forest. *Pychol Methods*, 14(4), 323-348.

Tartaglione, R. (1997) *Grammatica Italiana. Regole ed Esempi*. Firenze: Alma Edizioni.

Tauroza, S. & Allison, D. (1990) Speech rates in British English. *Applied Linguistics,* 11, 90-105.

Tedeschi, R. (2007) Referring expressions in early Italian. A study on the use of lexical objects, pronouns and null objects in Italian pre-school children. In S. Baauw, J. van Kampen & M. Pinto (eds) *The Acquisition of Romance Languages. Selected Papers from The Romance Turn II* (pp. 201-216). Utrecht: LOT (LOT Occasional Series 8).

Thomas, M. (1993) *Knowledge of Reflexives in a Second Language*. Amsterdam and Philadelphia: John Benjamins.

Tomasello, M. (1992) *First Verb: A Case Study of Early Grammatical development*. Cambridge: Cambridge University Press.

Tomasello, M. (2000a) The item-based nature of children's early syntactic development. *Trends in Cognitive Science*, 4(4), 156-163.

Tomasello, M. (2000b) Do young children have adult syntactic competence? *Cognition*, 74, 209-253.

Tomasello, M. (2003) *Constructing a language. A usage-based theory of language acquisition*. Cambridge, Massachusetts and London: Harvard University Press.

Tomasello, M. (2005) Beyond formalities: The case of language acquisition. *The Linguistc Review*, 22, 209-253.

Tomasello, M. (2006a). Construction Grammar for Kids. Constructions SV1–SV11, 1–23.

Tomasello, M. (2006b) Acquiring Linguistic Constructions. In D. Kuhn & R.S. Siegler RS (eds) *Cognition, perception, and language* (*6th ed.)*. In W. Damon & R.M. Lerner (eds) Handbook *of child psychology*, vol. 2 (pp. 255-298). Hoboken, NJ: Wiley

Tomasello, M. & Abbot-Smith, K. (2002) A tale of two theories: response to Fisher. *Cognition* 83, 207-214.

Tomasello, M., Akhtar, N., Dodson K. & Rekau, L. (1997) Differential productivity in young children's use of nouns and verbs. *Journal of Child Langauge*, 24, 373-387.

Tomasello M. & Kruger, A. (1992) Joint attention on action: Acquiring verbs in non-ostensive contexts. *Journal of Child Language*, 19, 311-333.

Tomasello, M. & Olguin, R. (1993) Twenty-Three-Month-Old Children Have a Grammatical Category of Noun. *Cognitive Development*, 8, 452-464.

Tomasello, M. & Stahl, D. (2004) Sampling children's spontaneous speech: how much is enough? *Journal of Child Language*, 21, 101-121.

Tonelli, L., & Fabris, M. (2005). L'acquisizione della flessione verbale - esemplificazione di un metodo di ricerca. *AnnalSS*, 2, 13-34.

Wellman, H.M. (1990) *The Child's Theory of Mind*. Cambridge, Massachusetts and London: MIT Press.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523-541. DOI: 10.1111/j.1467-8624.2004.00691.x

Wilde-Astington, J. & Edward, M.J. (2010) The Development of Theory of Mind in Early Childhood. *Encyclopedia on Early Childhood Development* [online]. August 2016. URL: http://www.child-encyclopedia.com/social-cognition/according-experts/development-theory-mind-early-childhood. Accessed March 2017.

Yang, C. (2009) Who's Afraid of George Kingsley Zipfl? Unpublished Manuscript: University of Pennsylvania. Retrieved March 2014, from http://www.ling.upenn.edu/~ycharles/papers/zipfnew.pdf.

Yang, C. & Roeper, T. (2011) Minimalism and Language Acquisition. In C. Boeckx (ed.) *The Oxford Handbook of Linguistic Minimalism*. Oxford University Press. Retrieved Sep. 2016, from http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199549368.001.0001/oxfordhb-9780199549368-e-024

Yo, W., Ference, B.A, Cote, M.L & Schwartz, A. (2012) A Comparison of Logistic Regression, Logic Regression, Classification Tree, and Random Forests to Identify Effective Gene-Gene and Gene-Environmental Interactions. *Int J Appl Sci Technol*., 2012 August ; 2(7). Retrieved at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3686280/pdf/nihms458280.pdf

Zipf, G. K. (1949). *Human behaviour and the principle of least effort: An introduction to human ecology*. Cambridge, Massachusetts: Addison-Wesley.

2017

Northumbria
University

Luca Miorelli

THE DEVELOPMENT OF MORPHO-SYNTACTIC COMPETENCE IN ITALIAN-SPEAKING CHILDREN: A USAGE-BASED APPROACH

Volume II of II

Supervisors:
Prof. Ewa Dabrowska
Dr James Street

The Development of Morpho-Syntactic Competence in Italian-Speaking Children: A Usage-Based Approach

Luca Miorelli

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle upon Tyne for the degree of Doctor of Philosophy

Research undertaken in the Faculty of Arts, Design & Social Sciences

October 2017

TABLE OF CONTENTS

VOLUME II

Part V: Appendix

Page intentionally left blank

The Development of Morpho-Syntactic Competence in Italian-Speaking Children: A Usage-Based Approach

Volume II

Page intentionally left blank

Part V:

Appendix

Page intentionally left blank

19.

APPENDIX_I:

Figures, Tables and Examples

Page intentionally left blank

19.0. INTRODUCTION

Table 0.1: Constructions, form-meaning pairings. Elements in small letters are fully lexically-specific concrete words, whereas CAPITALS indicate SLOTS (that is, more schematic, lexically-unspecified semantic/functional generalisations).

	SYMBOLIC UNITS		
	simple	complex	
fully concrete	banana	kicked	I make those
partially schematic	b.........	PROCESS-ed	MAKER make THING
fully schematic	THING	PROCESS-INFLECTION	AGENT-PROCESS-PATIENT

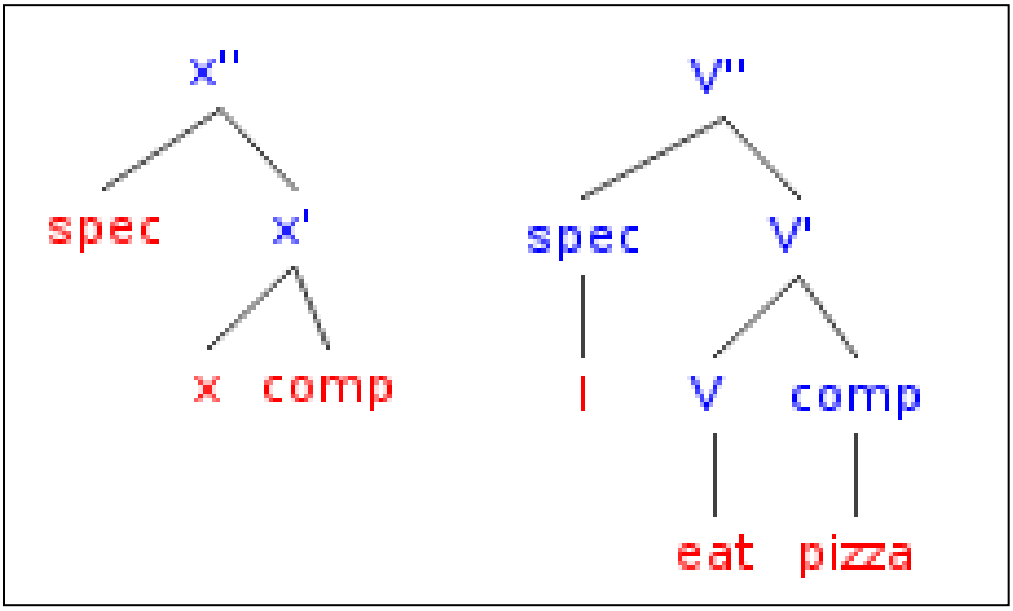


Figure 0.1: the x-bar structure.

1. The boy is angry.
2. Is the boy _ angry?
3. The boy who is screaming is angry.
4. *Is the boy who _ screaming is angry?
5. [The boy who is screaming] [is] [angry].
6. Is the boy who is screaming _ angry?

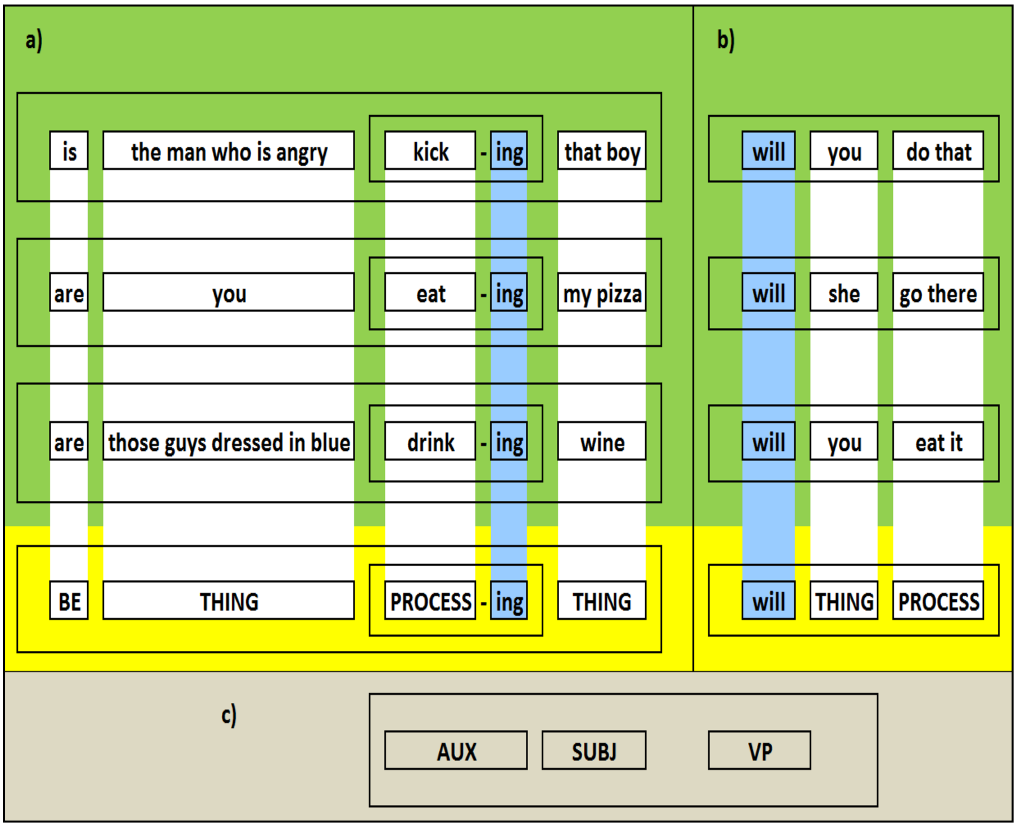


Figure 0.2: A UBA to learning syntactic questions in English. Concrete expressions from which schemas are inferred are in the green strip. Schemas are in the yellow strip. The grey strip (c) indicates that semantic generalisations (yellow strip) may gradually develop into more adultlike (possibly syntactic) ones. Slot formation (generalisations) is highlighted in white. Recurring lexical material is highlighted in blue.

19.1. Part I: BACKGROUND

19.1.1. THE ITALIAN LANGUAGE

Table 1.1: Types of adjectives in Italian.

type of adjective	SG M	SG F	PL M	PL F
only SG vs PL distinction	grand-e big-SG		grand-i big-PL	
SG vs PL and MASC vs FEM distinction	bell-o nice-M.SG	bell-a nice-F.SG	bell-i nice-M.PL	bell-e nice-F.PL

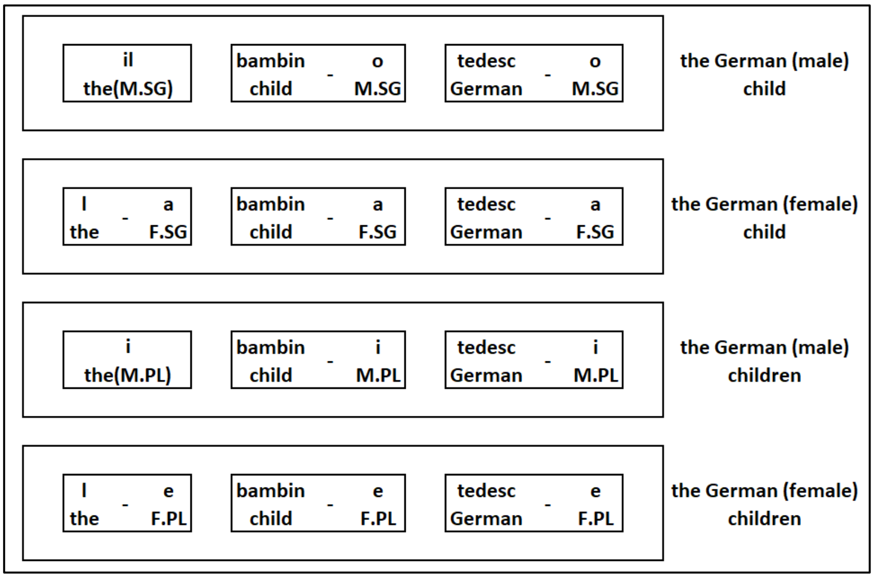


Figure 1.1: article-adjective-noun agreement in Italian and the regular gender-number markers (on adjectives and nouns) *-o*(M.SG), *-i*(M.PL), *-a*(F.SG), *-e*(F.PL).

1. a) *Una lama affilata*
   un-a *lam-a* *affilat-a*
   a-F.SG blade-F.SG sharp-F.SG

   b) *?? Un' affilata lama*
   ?? un *affilat-a* *lam-a*
   ?? a sharp-F.SG blade-F.SG

   "A sharp blade."

2. a) *I miei vecchi cappelli*
   i *mi-ei* *vecch-i* *cappell-i*
   the(M.PL) my-M.PL old-M.PL hat-M.PL
   "My old hats (my hats, which are old)."

   b) *I miei cappelli vecchi*
   i *mi-ei* *cappell-i* *vecch-i*
   the(M.PL) my-M.PL hat-M.PL old-M.PL
   "My old hats (the sub-class of my hats that are old)."

3. a)* *Una molto bella ragazza*

*	*un-a*	*molto*	*bell-a*	*ragazz-a*
*	a-F.SG	very	pretty-F.SG	girl-F.SG

b) *Una ragazza molto bella*

un-a	*ragazz-a*	*molto*	*bell-a*
a-F.SG	girl-F.SG	very	pretty-F.SG

"A very pretty girl."

4. a) *Una ragazza bellissima*

un-a	*ragazz-a*	*bell-issim-a*
a-F.SG	girl-F.SG	pretty-very-F.SG

b) *Una bellissima ragazza*

un-a	*bell-issim-a*	*ragazz-a*
a-F.SG	pretty-very-F.SG	girl-F.SG

"A gorgeous girl."

5. a) *Suo fratello*

su-o	*fratell-o*
his/her-M.SG	brother-M.SG

b) **Sua fratello*

su-a	*fratell-o*
his/her-F.SG	brother-M.SG

"Her brother."

6. a) *Il tavolino*

il	*tavol-in-o*
the(M.SG)	table-little(DIM)-M.SG

b) *Il piccolo tavolo*

il	*piccol-o*	*tavol-o*
the(M.SG)	little-M.SG	table-M.SG

c) *Il tavolo piccolo*

il	*tavol-o*	*piccol-o*
the(M.SG)	table-M.SG	little-M.SG

"The little table."

Table 1.2: the three conjugational classes of Italian.

Conjugation	Tematic Vowel TV	Example	Proportion of irregular verbs
1st Conjugation (or Conjugation I)	a	am - a - re love - TV - INF	four verbs and their derivatives
2nd Conjugation (or Conjugation II)	e	tem - e - re fear - TV - INF	95%
3rd Conjugation (or Conjugation III)	i	sent - i - re hear - TV - INF	10%

Table 1.3: the distribution of first, second and third conjugation verbs in Italian, according to different sources.

Primary sources used for this Research	Secondary sources on which primary sources draw	no. of verbs in each conjugation		
		Conjugation I	Conjugation II	Conjugation III
Barbieri (1971)	*Nuovo Vocabolario Zingarelli*	13,465	1,556	1,966
Orsolini & Marslen-Wilson (1997)	Dressler & Thornton (1991)	more than 3000	less than 400	more than 500
Say & Clahsen (2002)	De Mauro et al. (1993) and Bortolini et al. (1971)	1,709	403	238

Table 1.4: Conjugating verbs in Italian: bare root + aspect-mood-tense-person-number affixes.

Verb	Tense, mood and aspect	Conj	root	aspect-mood-tense-person-number affix	resulting form	meaning
amare (to love)	ind prs	I	am	o	amo	I love
amare (to love)	ind prs	I	am	ano	amano	they love
temere (to fear)	ind prs	II	tem	o	temo	I fear
temere (to fear)	ind prs	II	tem	ono	temono	they fear
sentire (to hear)	ind prs	III	sent	o	sento	I hear
sentire (to hear)	ind prs	III	sent	ono	sentono	they hear
amare (to love)	subjunctive prs	I	am	i	ami	(that) she/he love
temere (to fear)	subjunctive prs	II	tem	a	tema	(that) she/he fear
sentire (to hear)	subjunctive prs	III	sent	a	senta	(that) she/he hear
amare (to love)	imp	I	am	a	ama!	(you) love!
temere (to fear)	imp	II	tem	i	temi!	(you) fear!
sentire (to hear)	imp	III	sent	i	senti!	(you) hear!

Table 1.5: Conjugating verbs in Italian: stem + aspect-mood-tense-person-number affixes.

Verb	Tense, mood and aspect	Conj	stem (root + tv)	aspect-mood-tense-person-number affix	resulting form	meaning
amare (to love)	past-definite (or preterite)	I	am + a	sti	amasti	you loved
amare (to love)	past-definite (or preterite)	I	am + a	stress	amò	she/he loved
temere (to fear)	past-definite (or preterite)	II	tem + e	sti	temesti	you feared
temere (to fear)	past-definite (or preterite)	II	tem + e	stress	temè	she feared
sentire (to hear)	past-definite (or preterite)	II	sent + i	sti	sentisti	you heard
sentire (to hear)	past-definite (or preterite)	III	sent + i	stress	sentì	she/he heard
amare (to love)	inf	I	am + a	re	amare	to love
temere (to fear)	inf	II	tem + e	re	temere	to fear
sentire (to hear)	inf	III	sent + i	re	sentire	to hear

Table 1.6: Conjugating verbs in Italian: stem + aspect-mood-tense affixes + person-number affixes.

Verb	Tense, mood and aspect	Conj	stem (root + tv)	apsect-mood-tense	person-number affix	resulting form	meaning
amare (to love)	ind imperf	I	am + a	v	o	amavo	I used to love
temere (to fear)	ind imperf	II	tem + e	v	o	temevo	I used to fear
sentire (to hear)	ind imperf	III	sent + i	v	o	sentivo	I used to hear

Table 1.7: Conjugating verbs in Italian: root + aspect-mood-tense affixess + person-number affixes.

Verb	Tense, mood and aspect	Conj	root	aspect-mood-tense specific infix	person-number affix	resulting form	meaning
amare (to love)	fut	I	am	er	emo	ameremo	we will love
temere (to fear)	fut	II	em	er	emo	temeremo	we will fear
sentire (to hear)	fut	III	sent	ir	emo	sentiremo	we will hear

Table 1.8: irregular verbs, suppletivism: present indicative of *andare* "to go".

person-number	stem/root	aspect-mood-tense-person-number	resulting form	meaning
1.SG	**vad**	**o**	**vado**	**I go**
2.SG	**va**	**i**	**vai**	**you go**
3.SG	**va**	**ø**	**va**	**s/he goes**
1.PL	and	iamo	andiamo	we go
2.PL	and	ate	andate	you go
3.PL	**va**	**nno**	**vanno**	**they go**

Table 1.9: indicative present and past-definite (preterite) of Conjugation II *root-change verbs.*

Verb	Tense, mood and aspect	root	aspect-mood-tense-person-number	resulting form	meaning
prendere (to take)	ind prs	pre**nd**	iamo	prend-iamo	we take
prendere (to take)	past-definite (peterite)	pre**s**	i	pres-i	I took
ridere (to laugh)	ind prs	ri**d**	iamo	rid-iamo	we laugh
ridere (to laugh)	past-definite (peterite)	ri**s**	i	risi	I laughed
mettere (to put)	ind prs	me**tt**	iamo	mett-iamo	we put
mettere (to put)	past-definite (peterite)	mi**s**	i	mis-i	I put
leggere (to read)	ind prs	le**gg**	iamo	legg-iamo	we read
leggere (to read)	past-definite (peterite)	le**ss**	i	less-i	I read

Table 1.10: No-root-change verbs; past-definite (preterite).

No-root-change Verbs (*vendere* "to sell")						
Verb		stem (root + tv)	tense-specific interfix	person-number marker	resulting form	meaning
default option	vendere (to sell)	vend + e	-	stress	vendé	s/he sold
non-default option	vendere (to sell)	vend + e	tt	e	vendette	s/he sold
default option	vendere (to sell)	vend + e	-	rono	venderono	they sold
non-default option	vendere (to sell)	vend + e	tt	ero	vendettero	they sold

Table 1.11: Forming regular participles in Italian.

Verb	Conj	root	TV	particple marker	number-gender suffix	resulting form	meaning
amare (to love)	I	am	a	t	o/i	ama-t-o/i	loved-M.SG/M.PL
amare (to love)	I	am	a	t	a/e	am-a-t-a/e	loved-F.SG/F.PL
temere (to fear)	II	tem	u*	t	o/i	tem-u-to/i	feared-M.SG/M.PL
temere (to fear)	II	tem	u*	t	a/e	tem-u-t-a/e	feared-F.SG/F.PL
sentire (to hear)	III	sent	i	t	o/i	sent-i-t-o/i	heard-M.SG/M.PL
sentire (to hear)	III	sent	i	t	a/e	senti-t-a/e	heard-F.SG/F.PL

* *-u* - is the allomorph of conjugation II's tv *–e-*.

Table 1.12: Past participle of irregular Conjugation II verb.

process	verb	regular root	modified root	-/st/ or -/t/	gender-number suffix (M.SG)	resulting form	meaning
same root of past-definite (preterite)	ridere	rid	ris	-	o	riso	laughed
other root + -/st/ or -/t/	piangere	piang	pian	t	o	pianto	cried

7. *Le ragazze sono cadute*
   *l-**e*** *ragazz-**e*** *sono* *cad-u-t-**e*** .
   the-**F.PL** girl-**F.PL** be(PRS.3.PL) fall-TV(conj.II)-PTCP-**F.PL**
   "The girls fell."

8. *Sara ha spostato la sedia*
   Sara *ha* *spost-a-t-**o***
   Sara has move-TV(conj.I)-PTCP-**M.SG(default)**
   *l-**a*** *sedi-**a***
   the-**F.SG** chair-**F.SG**
   "Sara moved the chair."

9. a) *Le ho mangiate*
   *l-**e*** *ho* *mangi-a-t-**e***
   clitic.3.ACC.-**F.PL** have(PRS.1.SG) eat-TV(conj.I)-PTCP-**F.PL**

   b) **Le ho mangiato*
   l-e*** *ho* *mangi-a-t-**o***
   *clitic.3.ACC-**F.PL** have(PRS.1.SG) eat-TV(conj.I)-PTCP-**M.SG**

   "I ate them."

10. *Nevica molto*
    nevic-a *molto*
    snow-PRS.3.SG a.lot
    "It snows a lot."

11. *Guarda! Non è stanca qui?*
guard-a! non é stanca-a qui?
look-IMP.2.SG! not is tired-F.SG here?
"Look! Isn't she tired, here?"
Context: while Looking at a girl's picture.
(Serratrice, 2005, p. 442).

12. *Vado via*
vad-o via
go-PRS.1.SG away.
"I go away."

13. a) *Cosa hanno fatto Chiara e Marco ieri?*
cosa hanno fatto Chiara e
what have(PRS.3.PL) done Chiara and
Marco ieri?
Marco yesterday?
"What did Chiara and Marco do yesterday?"

b) *Sono andati alla festa, ma lei è andata a casa presto*
sono and-a-t-i a=(l)l-a
be(PRS.3.PL) go-TV(conj.I)-PTCP-M.PL to=the-F.SG
fest-a, ma lei é anda-t-a
party-F.SG, but she is go-TV(conj.I)-PTCP-F.SG
a cas-a presto
to home-F.SG early
"They went to the party, but she went home early."

14. *Dopo la festa, Paolo mi ha dato un passaggio a casa*
dopo l-a fest-a, Paolo mi
after the-F.SG party-F.SG, Paolo clitic.1.SG.DAT
ha d-a-to un
has give-TV(conj.I)-PTCP a(M.SG)
passaggi-o a cas-a
lift-M.SG to home-F.SG
"After the party, Paolo gave me a lift home."
(From Serratrice, 2005, p. 444)

15. *IO, ho pulito dopo la festa (non tu)!*
IO, ho pul-i-to dopo
I, have(PRS.1.SG) clean-TV(conj.III)-PTCP after
l-a fest-a (non tu)!
the-F.SG party-F.SG (not you)!
"***I***, cleaned after the party (not you)!"

Table 1.13: the Italian pronominal system (no possessives). Adapted from Cordin and Calabrese (2001).

Type of Pronoun	case	Singular Forms				Plural Forms			
		1	2	3 masc	3 fem	1	2	3 masc	3 fem
TONIC (or Strong) PRONOUNS	nom	io	tu	lui/egli**	lei/ella**	noi	voi	loro/ essi**	loro/esse**
	acc	me	te	lui/esso**	lei/essa**	noi	voi	loro	loro
CLITIC PRONOUNS	acc	mi	ti	lo	la	ci	vi	li/gli	le
	dat	mi	ti	gli/ci	le/gli/ci	ci	vi	gli/ loro	gli/ loro

** formal language

16. a) *Dove sono **i coltelli**?*

dove	*sono*	***i***	***coltell-i***
where	be(PRS.3.PL)	**the(M.PL)**	**knife-M.PL**

"Where are the knives?"

b) ***Li** ho presi io*

l-i	*ho*	*pres-i*	*io*
clitic.3.ACC-M.PL	have(PRS.1.SG)	taken-M.PL	I

"I took them."

17. *Lo vedo*

l-o	*ved-o*
clitic.3.ACC-M.SG	see-PRS.1.SG

"I see it/him."

18. *Mangiarlo*[1]

*mangi-a-r=**l-o***
eat-TV(conj.I)-INF=**clitic.3.ACC-M.SG**
"To eat it/him."

19. *Mangialo!*

*mangi-a=**l-o**!*
eat-IMP.2.SG=**clitic.3.ACC-M.SG**
"Eat it/him!"

[1] According to Lehmann's (1982) guidelines, the symbol "=" indicates morpho-phonological merging (i.e. cliticisation). In (18) it indicates that the clitic merges with the verb.

20. a) *Non lo mangiare!*
non ***l-o*** *mangi-a-re!*
not **clitic.3.ACC-M.SG** eat-TV(conj.I)-INF/IMP.2.SG

b) *Non mangiarlo!*
non *mangi-a-r=**l-o***
not eat-TV(conj.I)-INF/IMP.2.SG=**clitic.3.ACC-M.SG**

"Don't eat it/him!"

21. *Mi ha dato il tuo libro*
mi *ha* *d-a-to* *un*
clitic.1.SG.DAT has give-TV(conj.I)-PTCP a(M.SG)
libr-o
book-M.SG.
"(S/he) gave me a book."

22. a) *Marco vuole mangiarlo!*
Marco *vuol-e* *mangi-a-r=**l-o***
Marco want-PRS.3.SG eat-TV(conj.I)-INF=**clitic.3.ACC-M.SG**

b) *Marco lo vuole mangiare*
Marco ***l-o*** *vuol-e*
Marco **clitic.3.ACC-M.SG** want-PRS.3.SG
mangi-a-re
eat-TV(conj.I)-INF.

"Marco wants to eat it/him."

(Cordin & Calabrese, 2001, p. 587)

23. a) *Carlo vuole che tu lo prenda*
Carlo *vuol-e* *che* *tu*
Carlo want-PRS.3.SG that you(2.SG.NOM)
l-o *prend-a.*
clitic.3.ACC-M.SG take-SBJV.PRS.2.SG

b) * *Carlo lo vuole che tu prenda*
Carlo ***l-o*** *vuol-e* *che*
Carlo **clitic.3.ACC-M.SG** want-PRS.3.SG that
tu *prend-a*
you(2.SG.NOM) take-SBJV.PRS.2.SG

"Carlo wants you to take it/him."

(Cordin & Calabrese, 2001).

24. a) *Lo voglio disperatamente*

l-o	*vogli-o*	*disperata-mente*
clitic.3.ACC-M.SG	want-PRS.1.SG	desperate-ly

b)* *Lo disperatamente voglio*

****l-o***	*disperata-mente*	*vogli-o*
clitic.3.ACC-M.SG	desperate-ly	want-PRS.1.SG

"I desperately want it/him."

25. *Glielo dico*

gli=(e)=l-o	*dic-o*
clitic.3.DAT.M.SG=(e)=clitic.3.ACC-M.SG	say-PRS.1.SG

("I say him/it to him.")
"I say that to him."

26. a) *Sara ci ha vist**o***

Sara	*ci*	*ha*	*vis-t-**o***
Sara	clitic.1.PL.ACC	has	see-PTCP-**M.SG**

b) *Sara ci ha vist**e***

Sara	*ci*	*ha*	*vis-t-**e***
Sara	clitic.1.PL.ACC	has	see-PTCP-**F.PL**

"Sara saw us."
Context: *ci* "us" refers to two or more girls.

27. *Il topo*$_i$*, la volpe lo*$_i$ *mangia*

[il	*top-**o]**$_i$,*	*l-a*	*volp-e*
[the(**M.SG)**	mouse-**M.SG]**$_i$,	the-F.SG	fox(F.)-SG
*[l-**o**]*$_i$	*mangi-a*		
[clitic.3.ACC-**M.SG**]$_i$	eat-PRS.3.SG		

"The fox eats the mouse."

28. *Il tetto si è rovinato tutto*

il	*tett-o*	*si*	*è*
the(M.SG)	roof-M.SG	clitic.3.REFL	is
rovin-a-t-o	*tutt-o*		
ruin-TV(conj.I)-PTCP-M.SG	all-M.SG		

"The roof got all spoiled."

29. *Il signore si è sporcato tutto*

il	*signor-e*	*si*	*è*
the(M.SG)	gentleman(M.)-SG	clitic.3.REFL	is
sporc-a-t-o	*tutt-o*		
dirty-TV(conj.I)-PTCP-M.SG	all-M.SG		

"The gentleman completely dirtied himself" or "The gentleman got all dirty".

30. *Si è spesso ignorati dai politici*
si è spesso ignor-a-t-i
clitic.IMPRS.NOM is often ignore-TV(conj.I)-PTCP-M.PL
da=i politic-i
by=the(M.PL) politician-M.PL
“People are often ignored by politicians.”

31. *Si mangiano i pomodori*
si mangi-ano i pomodor-i
clitic.IMPRS.NOM eat-PRS.3.PL the(M.PL) tomato-M.PL
“Tomatoes get/are eaten.”

32. *Non si fa così*
non si fa così
not clitic.IMPRS.NOM does like.that
“That is not the way to do it” or “This is not to be done” or “That’s not the way to behave”.

33. *I RAGAZZI rincorrono il cane*
i RAGAZZ-I rincorr-ono il can-e
the(M.PL) boy-M.PL chase-PRS.3.PL the(M.SG) dog(M)-SG
“They boys chase the dog.”

34. *Lo ha cucinato la mamma*
l-o ha cucin-a-t-o
clitic.3.ACC-M.SG has cook-TV(conj.I)-PTCP-M.SG
l-a mamm-a
the-F.SG mum-F.SG
“Mum cooked him/it.”

35. LD
a) *A Marco, Sara dà un oracchiotto.*
a Marco, Sara dà un orsacchiott-o
to Marco, Sara gives a(M.SG) teddy.bear-M.SG.
“Sara gives Marco a teddy bear.”

b) *A Marco$_i$, Sara gli$_i$ dà un oracchiotto.*
[a Marco]$_i$ Sara [gli]$_i$ dà un
[to Marco]$_i$, Sara [clitic.3.DAT.M.SG]$_i$ gives a(M.SG)
orsacchiott-o
teddy.bear-M.SG
“Sara gives Marco a teddy bear.”

c) **Marco, Sara dà un oracchiotto.*
**Marco, Sara dà un orsacchiott-o .*
Marco, Sara gives a(M.SG) teddy.bear-M.SG
“*Sara gives a teddy bear Marco.”

36. FOC

a) *A MARCO, Sara dà un oracchiotto.*
A MARCO, Sara dà un orsacchiott-o
to Marco, Sara gives a(M.SG) teddy.bear-M.SG.
"Sara gives a teddy bear TO MARCO."

b)* *A MARCOi, Sara glii dà un oracchiotto.*
**[A MARCO]i Sara* [*gli*]*i*
[TO MARCO]*i*, Sara [clitic.3.DAT.M.SG]*i*
dà un orsacchiott-o
gives a(M.SG) teddy.bear-M.SG.
"Sara gives Marco a teddy bear."

37. HT

a) Marcoi, Sara glii dà un oracchiotto.
[Marco]i , Sara [gli]i dà
[Marco]*i*, , Sara [clitic.3.DAT.M.SG]*i* gives
un orsacchiott-o
a(M.SG) teddy.bear-M.SG.
"Sara gives Marco a teddy bear."

b) *Marcoi, Sara dà un orsacchiotto a quel bambinoi*
[Marco]i, Sara dà un orsacchiott-o
[Marco]*i*, Sara gives a(M.SG) teddy.bear-M.SG
[a quel bambin-o]i
[to that(M.SG) child-M.SG]*i*
"Marco, Sara gives a teddy-bear to that child."

c) **Marco, Sara dà un oracchiotto.*
Marco, Sara dà un orsacchiott-o .
Marco, Sara gives a(M.SG) teddy.bear-M.SG
"Sara gives Marco a teddy bear."

38. a) LD and HT

Marcoi, tutti loi incolpano

[Marco]i, tutt-i [l-o]i incolp-ano

[Marco]*i*, all-M.PL [clitic.3.ACC-M.SG]*i* blame-PRS.3.PL

"Everybody blames Marco."

b) HT

[Marco]i, tutti incolpano quell'uomo cattivoi

[Marco]i, tutt-i incolp-ano

[Marco]*i*, all-M.PL blame-PRS.3.PL

[quell' uom-o cattiv-o]i

[that(M.SG) man-M.SG mean-M.SG]*i*

"Marco, everybody blames that mean man."

*c) *Marco, tutti incolpano*

Marco, tutt-i incolp-ano

Marco, all-M.PL blame-PRS.3.PL

"Everybody blames Marco."

39. a) *IL GATTO, ho buttato fuori*

IL GATT-O, ho

the(M.SG) cat-M.SG, have(PRS.1.SG)

butt-a-to fuori

throw-TV(conj.I)-PTCP out

b)* *[IL GATTO]i, loi ho buttato fuori.*

** [IL GATT-O]i, [l-o]i*

* [the(M.SG) cat-M.SG]*i*, [clitic.3.ACC-M.SG]*i*

ho butt-a-t-o fuori

have(PRS.1.SG) throw-TV(conj.I)-PTCP-M.SG out

"I threw THE CAT out."

40. *(Lo)i porto domani, il dolcei*

([l-o] i) port-o domani,

([clitic.3.ACC-M.SG]*i*) bring-PRS.1.SG tomorrow,

[il dolc-e]i

[the(M.SG) dessert(M.)-SG]*i*

"Tomorrow I'll bring the dessert."

(Benincà with al., 2001, pp. 160-161).

41. a. *Quando ha detto, Giorgio, che avrebbe smesso di piovere, a voi?*
quando ha detto, Giorgio, che avr-ebbe
when has said, Giorgio, that have-COND.PRS.3.SG
smesso di piov-e-re, a voi
stopped to rain-TV(conj.II)-INF, to you(2.PL)

b. *Quando ha detto, Giorgio, a voi, che avrebbe smesso di piovere?*
quando ha detto, Giorgio, a voi, che
when has said, Giorgio, to you(2.PL) that
avr-ebbe smesso di piov-e-re
have-COND.PRS.3.SG stopped to rain-TV(conj.II)-INF

c. *Quando ha detto, a voi, che avrebbe smesso di piovere, Giorgio?*
quando ha detto, a voi, che
when has said, to you(2.PL) that
avr-ebbe smesso di piov-e-re,
have-COND.PRS.3.SG stopped to rain-TV(conj.II)-INF ,
Giorgio
Giorgio

"When did Giorgio tell you that it would stop raining?"

42. *I pomodori piacciono a Marco*
i pomodor-i piacci-ono a Marco
the(M.PL) tomato-M.PL like-PRS.3.PL to Marco
"Marco likes tomatoes."

43. *A Marco piacciono i pomodori*
a Marco piacci-ono i pomodor-i
to Marco like-PRS.3.PL the(M.PL) tomato-M.PL
"Marco likes tomatoes."

44. *[A Marco]$_i$ gli$_i$ piacciono i pomodori*
[a Marco]$_i$ [gli]$_i$ piacci-ono
[to Marco]$_i$ [clitic.3.DAT.M.SG]$_i$ like-PRS.3.PL
i pomodor-i
the(M.PL) tomato-M.PL
"Marco likes tomatoes."

45. *A me, mi piacciono i pomodori.*
[a me]$_i$ [mi]$_i$ piacci-ono i
[to me]$_i$ [clitic.1.SG.DAT]$_i$ like-PRS.3.PL the(M.PL)
pomodor-i
tomato-M.PL
"I like tomatoes."

46. *Rincorrono la gallina, le volpi*

rincorr-ono	*l-a*	*gallin-a,*	*l-e*	*volp-i*
chase-PRS.3.PL	the-F.SG	chicken-F.SG,	the-F.PL	fox(F.)-PL

"The foxes chase the chicken."

47. *LA CASA, ha pulito Marco*

L-A	*CAS-A,*	*ha*
the-F.SG	house-F.SG,	has

pul-i-to	*Marco*
clean-TV(conj.III)-PTCP	Marco

"Marco has cleaned THE HOUSE (not something else)."

19.1.2. A USAGE-BASED APPROACH TO LANGUAGE

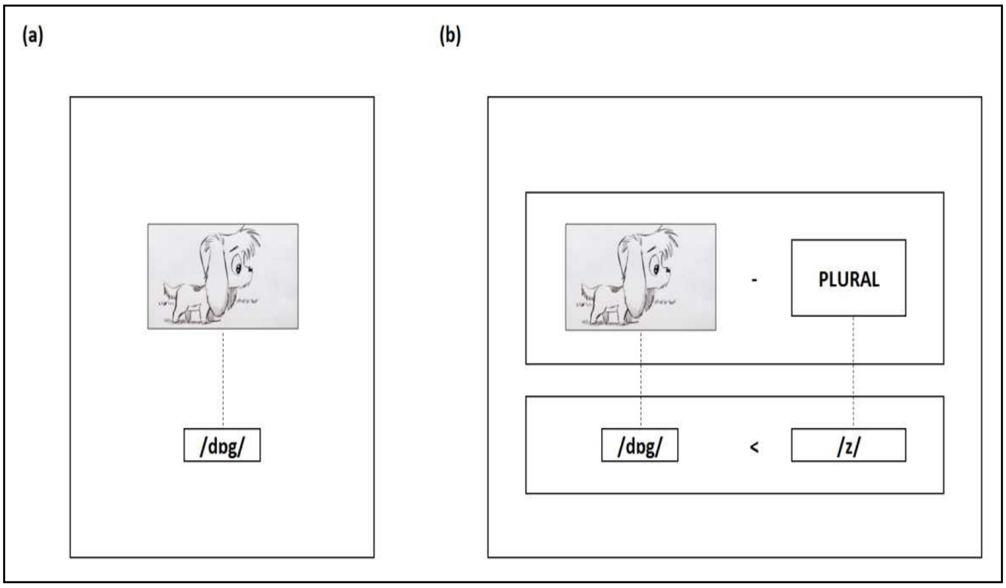


Figure 2.1: Graphic representation of symbolic units. The bottom row represents the phonological pole and the upper row represents the semantic pole. Dotted lines represent relationships of symbolisation between the two poles. The hyphen (-) stands for semantic integration and the symbol "<" stands for linear order (or temporal sequence).

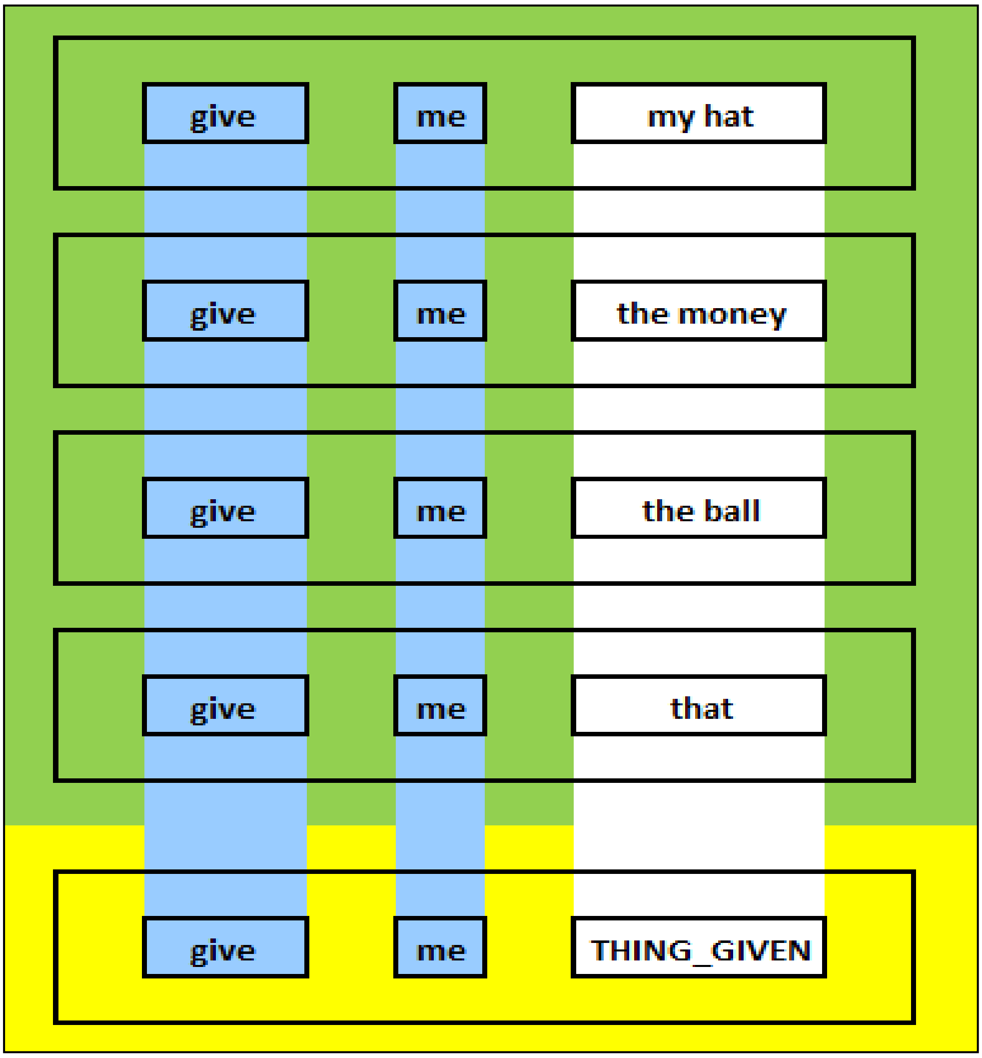


Figure 2.2: abstracting the schematic unit *give-me-THING_GIVEN* (yellow strip) from concrete expressions (green strip). The component parts shared by the schema (yellow strip) and its instantiations (green strip) are highlighted in blue. The variable elements across which *THING_GIVEN* represents a schematisation are highlighted in white.

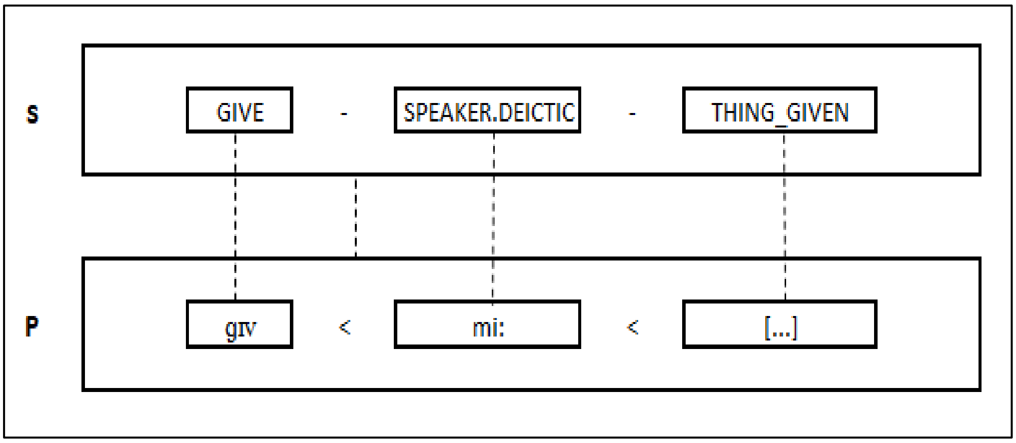


Figure 2.3: the schematic unit *give-me-THING_GIVEN*. The top line represents the semantic pole (S), where hyphens (-) indicate semantic integration. The bottom line represents the phonological pole (P), where the symbol "<" represents linear order.

Table 2.1: Symbolic Units. CAPITAL LETTERS indicate slots, that is, generalisations that represent phonologically-unspecified elements of a (schematic) symbolic unit. *Small letters* indicate elements whose semantic pole is fully specified at the phonological pole (refer back to fig. 2.3). The unit [b…..] indicates words whose initial morpheme is /b/. The psychological reality of such a unit is confirmed by the fact that we can engage in games in which we think of words whose initial phoneme is /b/. Such a unit is unspecified at the semantic pole. Yet, it is partially schematic because part of its phonological pole is (partially) specified.

SYMBOLIC UNITS			
	simple	complex	
fully concrete	banana	kicked	I make those
partially schematic	b..........	PROCESS-ed	MAKER make THING
fully schematic	THING	PROCESS-INFLECTION	AGENT-PROCESS-PATIENT

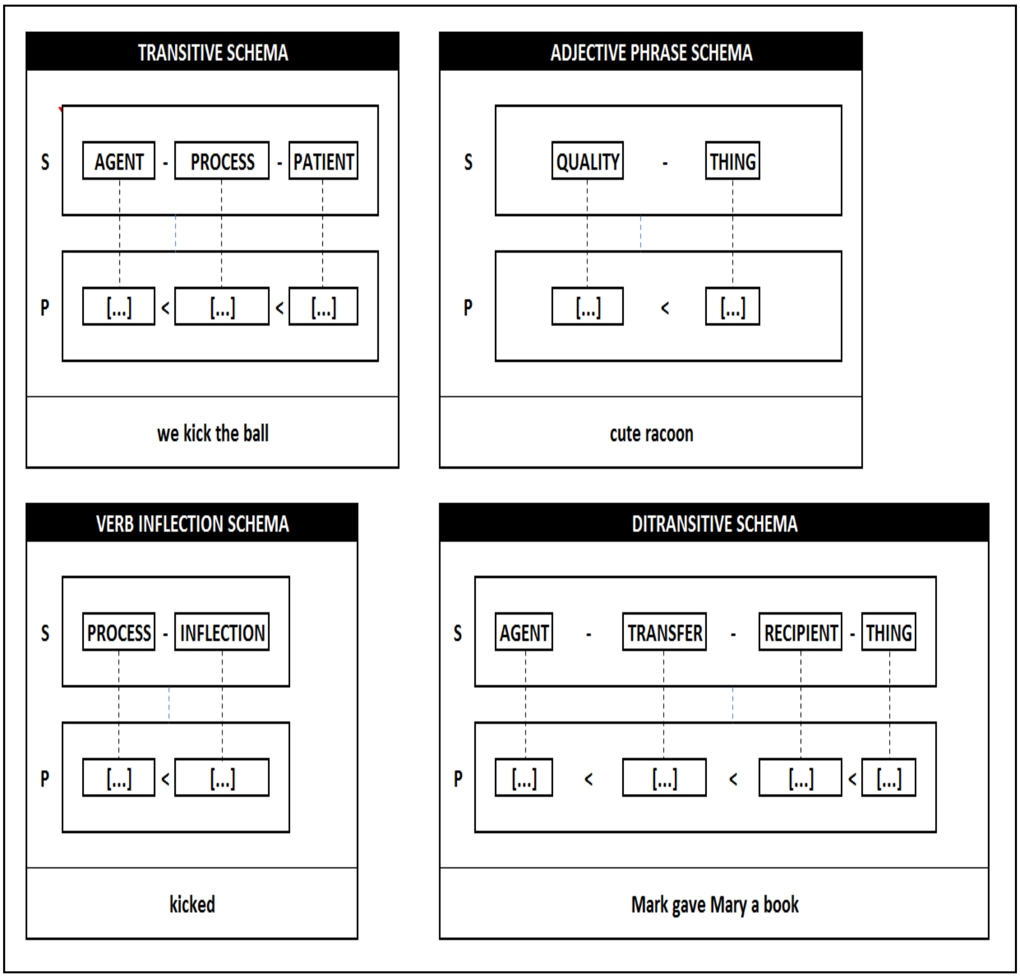


Figure 2.4: constructional schemas. For each schema, its semantic pole (S), its phonological pole (P) and a concrete instantiation of it are provided. Dashed lines indicate relationships of symbolisation. The symbol "<" indicates linear sequence at P. The hyphen (–) indicates integration at S.

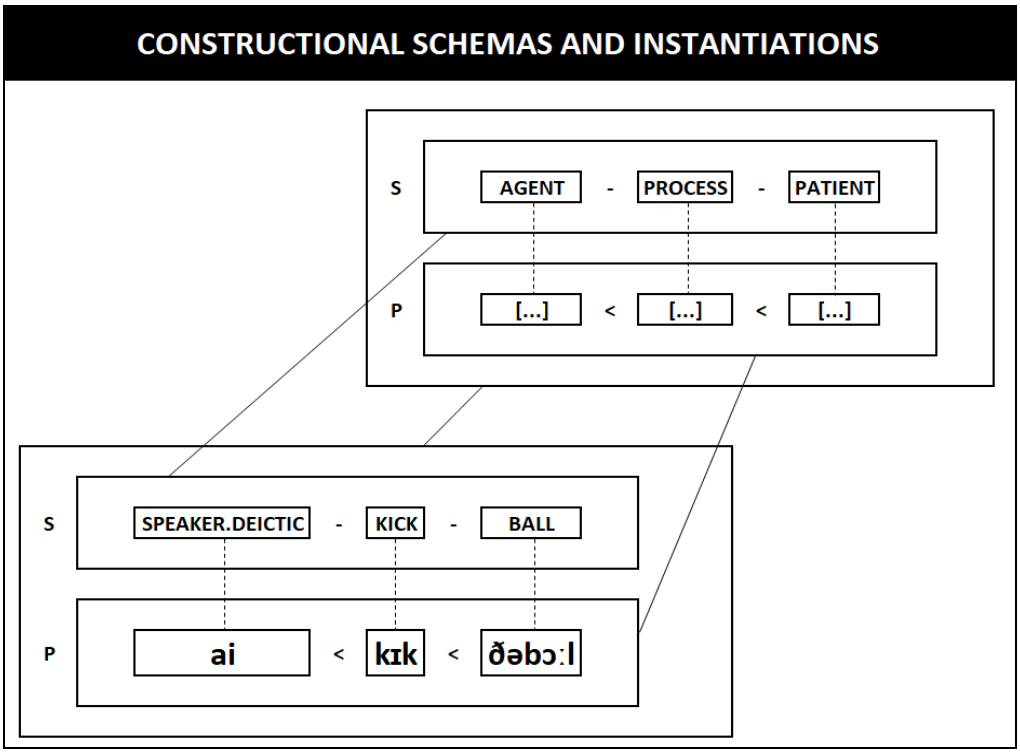


Figure 2.5 The semantic (S) and phonological (P) poles of *AGENT-PROCESS-PATIENT* and *I kick the ball*. Dashed lines represent relationships of symbolisation. Continuous lines indicate relationships of elaboration/instantiation.

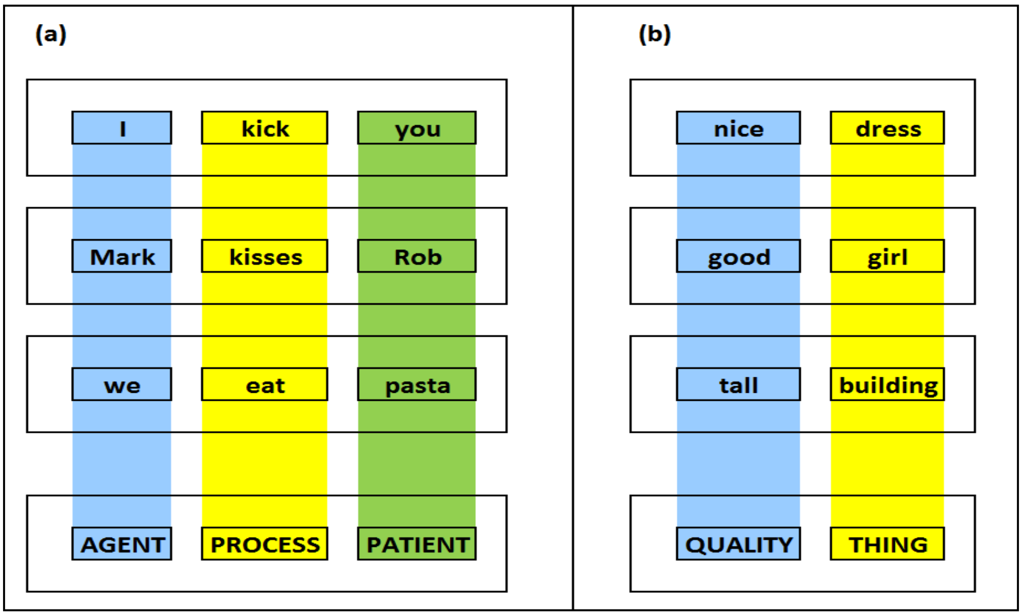


Figure 2.6: Abstracting constructional schemas from their concrete instantiations.

1. Mark kicked Rob the ball.

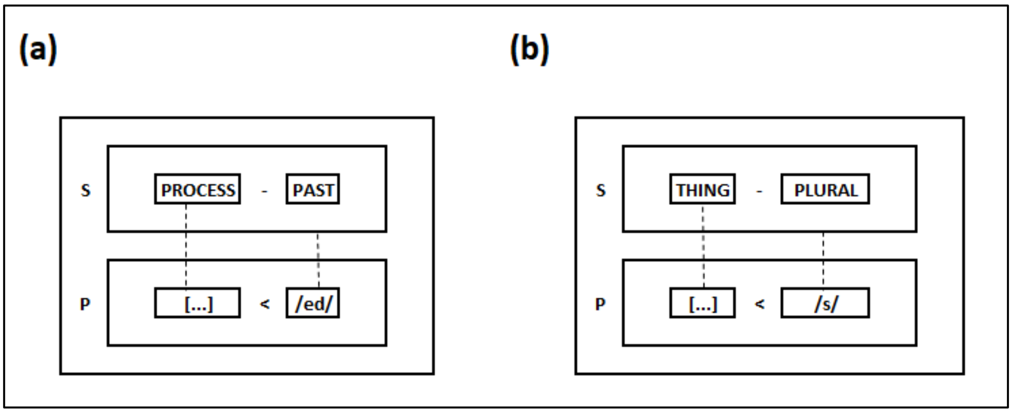


Figure 2.7: bound morphemes as lexically-bound schemas. The upper line symbolises the semantic pole (S), whereas the phonological pole (P) is reported on the bottom line. Dashed lines represent relationships of symbolisation between the two poles. Hyphens represent integration at the semantic pole, whereas the symbol "<" represents temporal sequence at the phonological pole.

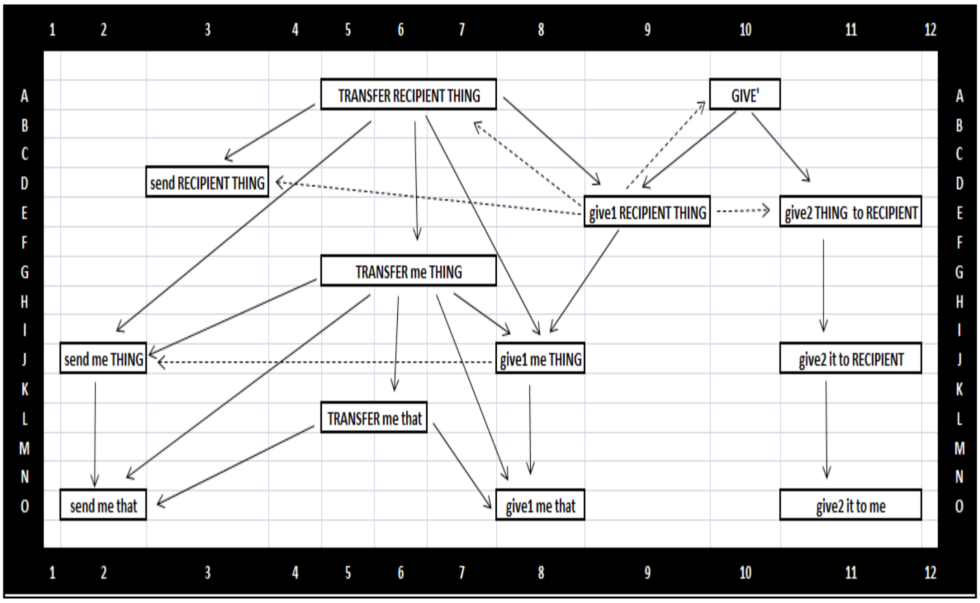


Figure 2.8: network of symbolic units (based on Langacker, 2000, 2008). Symbolic units are enclosed in rectangles. *Small letters* indicate elements specified at the phonological pole. *CAPITAL LETTERS* indicate slots. Dashed arrows indicate relationships of extension. Solid arrows indicate relationships of elaboration.

2. *Ho letto un libro*
ho *l-e-tto*
have(PRS.1.SG) read-TV(conj.II)-PTCP
un *libr-o*
a(M.SG) book-M.SG
"(I) have read a book."

3. *UN LIBRO, ho letto*
UN *LIBR-O,* *ho* *l-e-tto*
a(M.SG) book-M.SG have(PRS.1.SG) read-TV(conj.II)-PTCP
"A BOOK, (I) have read."

4. **Mangiai lo*
   mangi-a-i *l-o*
   eat-TV(conj.I)-PST.1.SG clitic.3.ACC-M.SG
   "I ate it."

5. *Lo mangiai*
   l-o *mangi-a-i*
   clitic.3.ACC-M.SG eat-TV(conj.I)-PST.1.SG
   "I ate it."

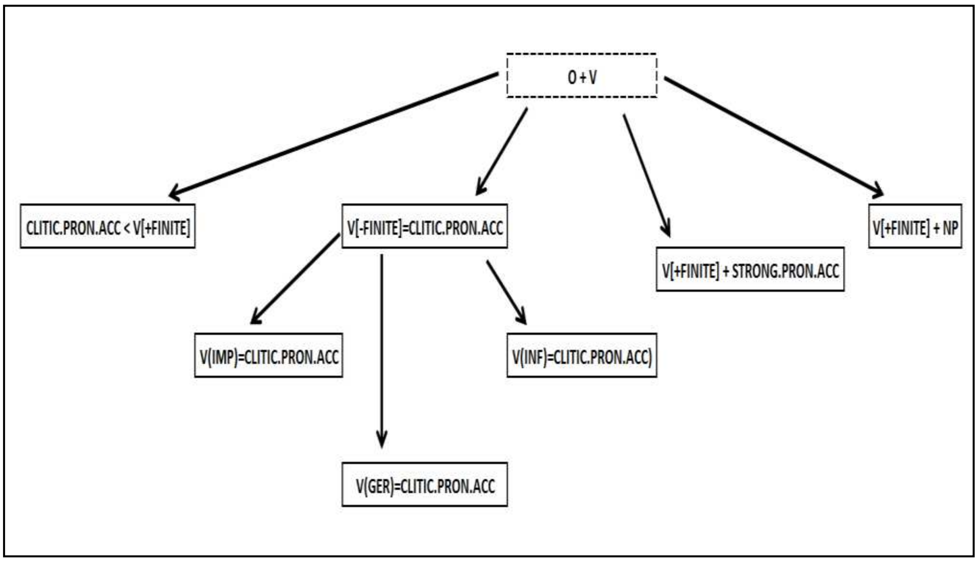


Figure 2.9: Verb and direct object in Italian. The symbol "<" indicates compulsory linear order. The symbol "=" indicates morpho-phonological merging. The symbol "+" indicates that the linear order is free (not specified). The dashed rectangle in which *O+V* is enclosed indicates that the highest-level schema is unlikely to be an entrenched unit available to sanction linguistic expressions. Arrows indicate relationships of elaboration/instantiation.

6. Give me that, now.

7. Now, give me that.

8. I need a chair.

9. *NEEDER-need-THING_NEEDED.*

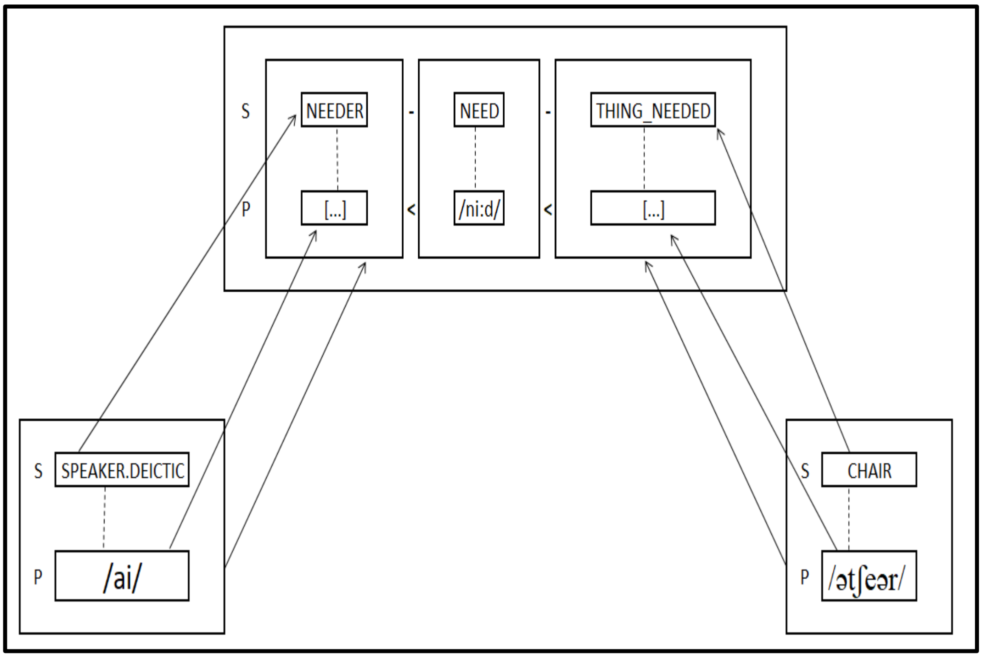


Figure 2.10 Superimposition: dashed lines indicate relationships of symbolisation between semantic (S) and phonological (P) pole. Arrows indicate that the more specific units elaborate the more schematic ones at semantic, phonological and symbolic levels.

.

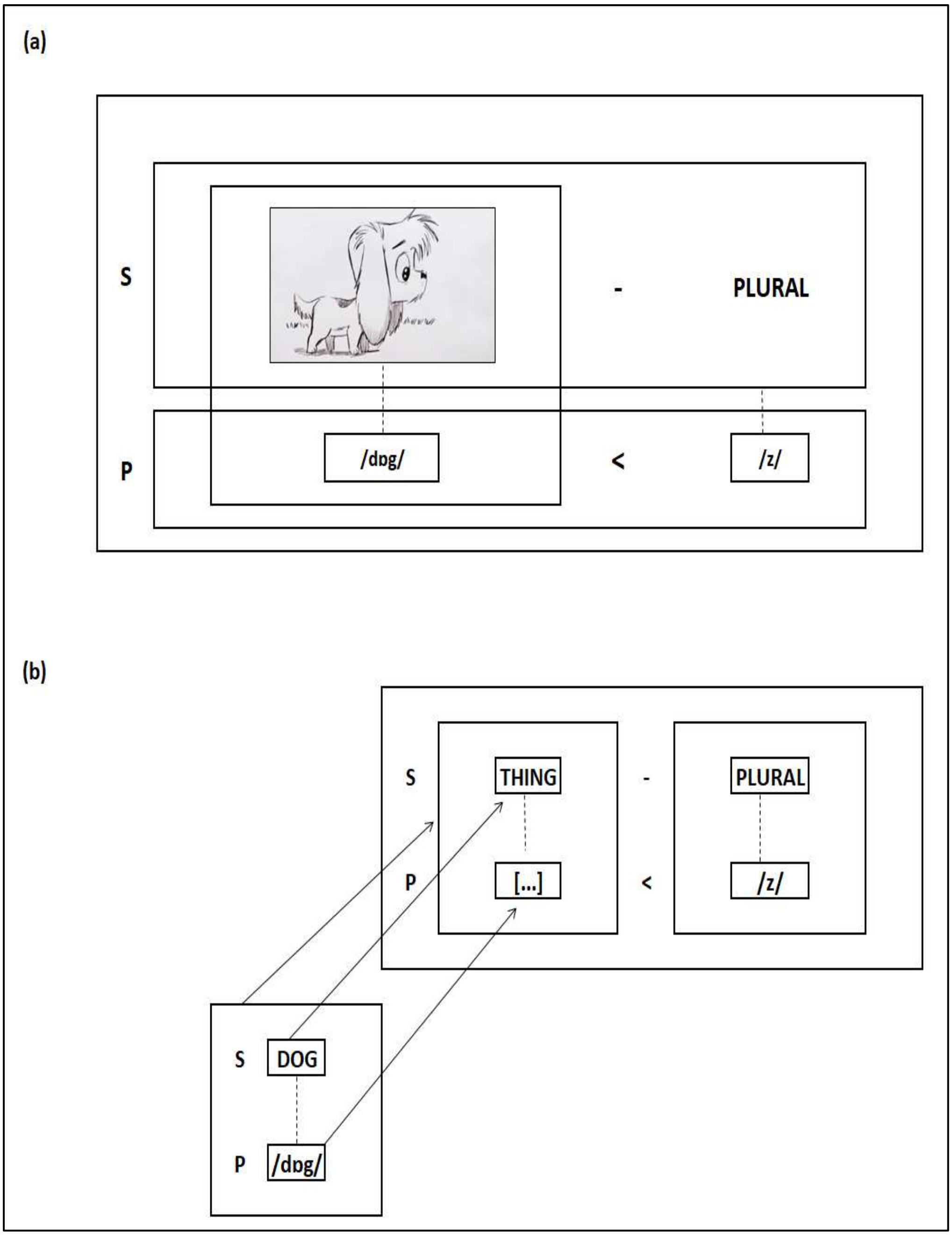


Figure 2.11: ***dogs*** **derived by superimposing** ***dog*** **and** ***THINGs*****. Arrows indicate relationships of elaboration (from the filler to the schema). Dashed lines represent symbolic relationships between phonological pole (P) and semantic pole (S). The hyphen (-) represents semantic integration and the symbol "<" indicates linear order.**

10. I don't know how to mum.

11. **badly boy.*

12. eat them!

13. Eat-THING_EATEN.

14. PROCESS-them.

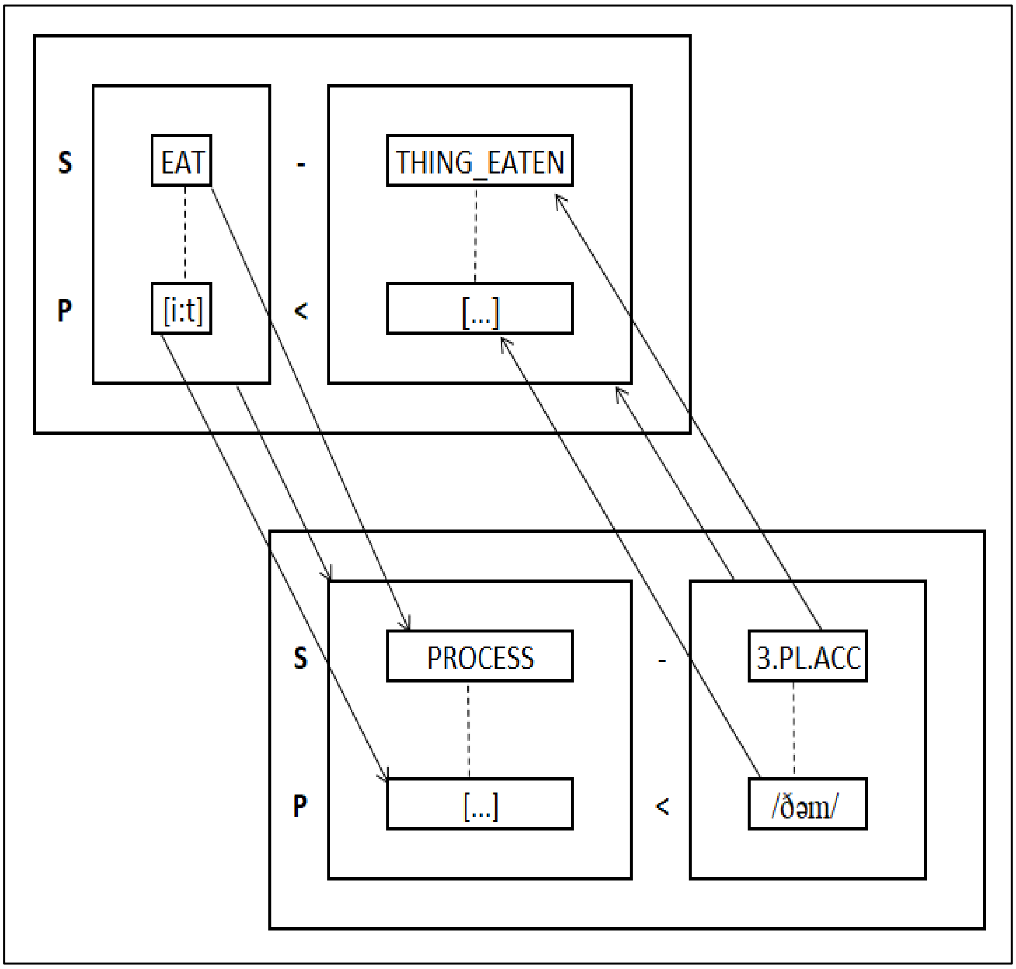


Figure 2.12 Mutual superimposing of two lexically bound schemas. Solid arrows indicate relationships of elaboration (from the more specific unit to the more schematic one). Dashed lines indicate relationships of symbolisation between semantic pole (S) and phonological pole (P). Hyphens (-) represent semantic integration and the symbol "<" indicates linear order.

15. Where are you hiding?

16. Where are you PROCESS-ing?

17. Where are you going?

18. Going.

19. WH are you PROCESS-ing?

20. Hiding.

21. Eat-THING_EATEN.

22. PROCESS-them.

23. Them.

24. What are you kicking?

25. Kick-KICKEE.

26. What-are-you-PROCESSing?

27. You are kicking what?

28. What are you kicking?

29. Eat them, now!

30. PROCESS-them.

31. Eat-THING_EATEN.

32. Now.

Table 2.2: the ditransitive construction: prototypical and extended meanings (based on Goldberg, 1995, pp. 38 and 72)

THE DITRANSITIVE CONSTRUCTION				
	PARAPHRASE	types of verbs that can appear in the construction	example	(putative) constructional schema
CENTRAL MEANING (prototype)	X causes Y to receive Z	verbs of giving, continous causation, instantaneous causation and ballistic motion	Joe gave Sally the ball	TRANFER NML NML
EXTENDED MEANING	X causes Y to receive Z, if X complies with what asserted	verbs of giving associated with satisfaction	Joe promised Bob a car	SATISFY NML NML
EXTENDED MEANING	X enables Y to receive Z	verbs of permission	Joe permitted Chris an apple	ENABLE NML NML
EXTENDED MEANING	X causes Y not to receive Z	verbs of refusal	Joe refused Bob a cookie	REFUSE NML NML
EXTENDED MEANING	X intends to cause Y to receive Z	verbs of creation and obtaining	Joe baked Bob a cake	CREATION NML NML
EXTENDED MEANING	X acts to cause Y to receive Z at some future point in time	verbs of future transfer	Joe bequeathed Bob a fortune	FUTURE.TRANSFER NML NML

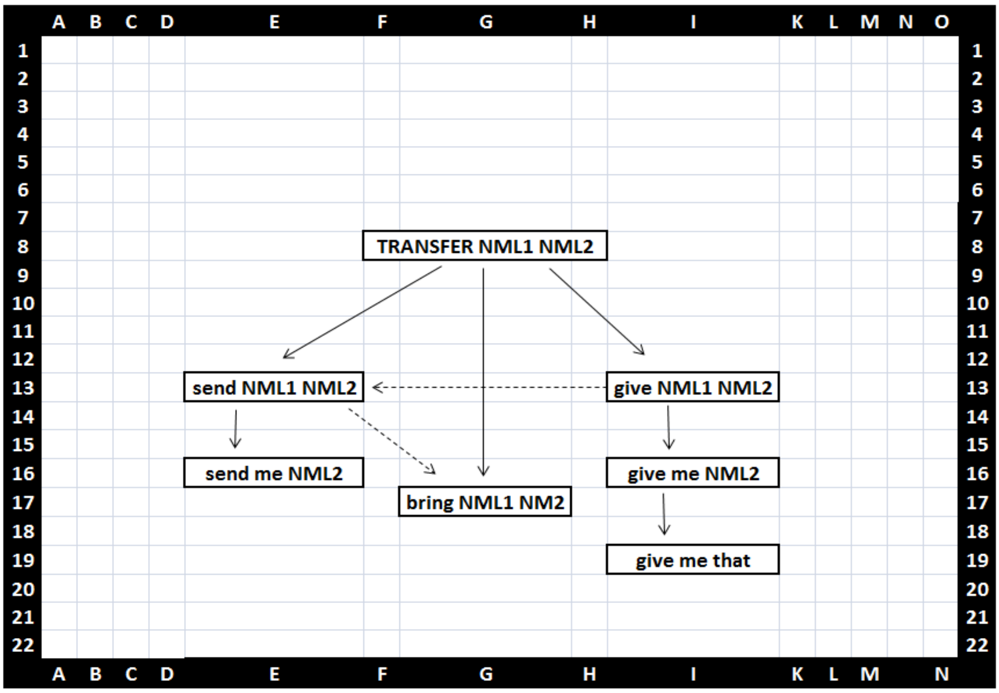


Figure 2.13: the ditransitive complex category 1 of 5: original nucleus. See fig. 2.8 on how to read this and the following figures (fig. 2.14-2.18) .

33. Rob baked Mary a cake.

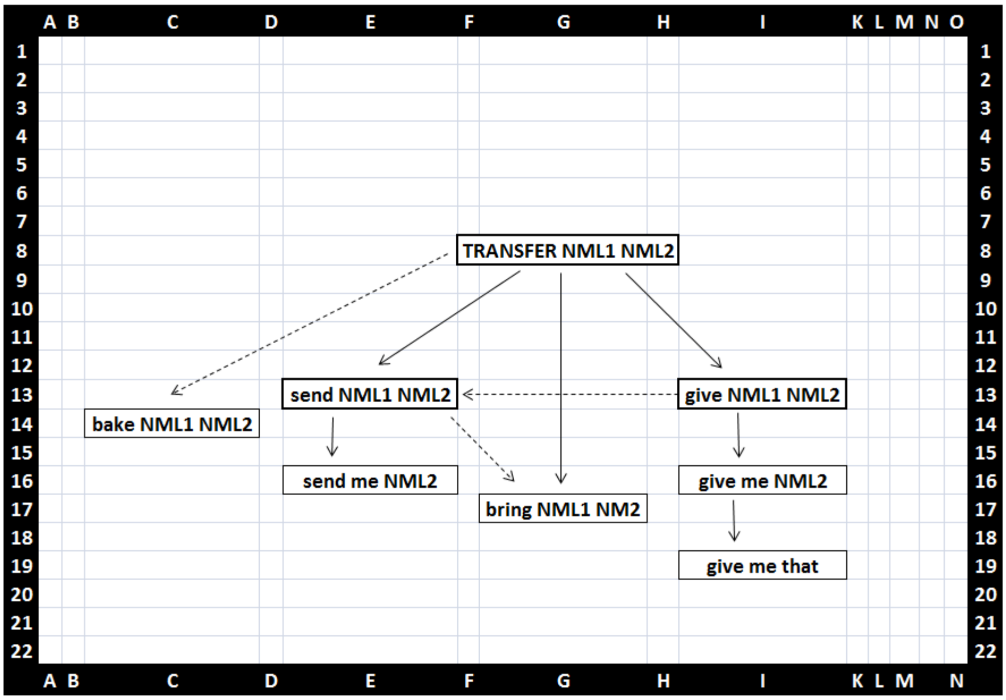


Figure 2.14: the ditransitive complex category; 2 of 5.

34. Mary wrote Mark a long letter.

35. [bake-NML1-NML2] - - ► Mary wrote Mark a long letter.

36. [TRANSFER-NML1-NML2] - - ► Mary wrote Mark a long letter.

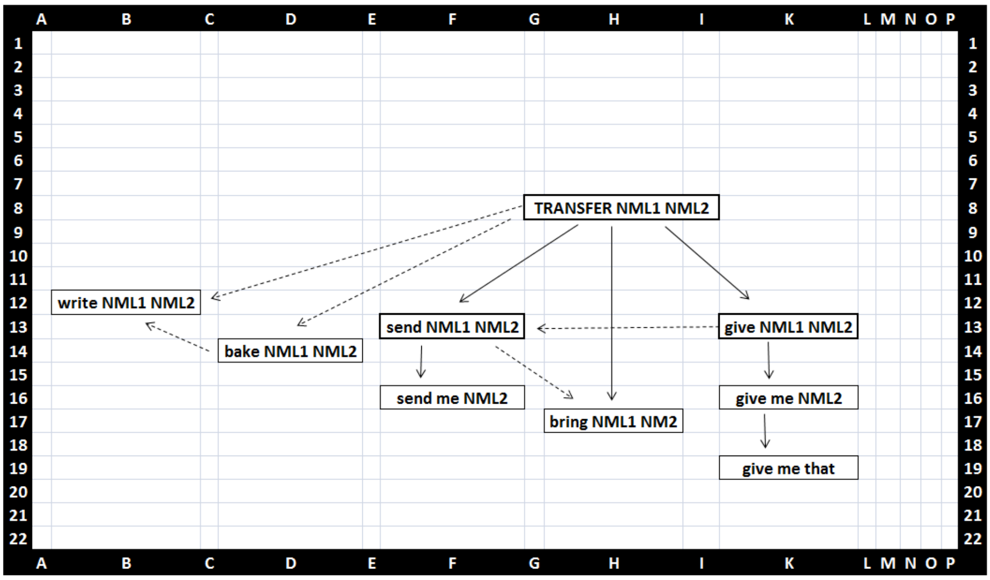


Figure 2.15: the ditransitive complex category; 3 of 5.

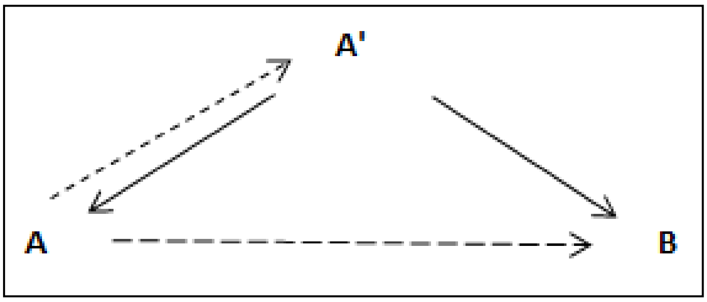


Figure 2.16: schematisation, instantiation and extension.

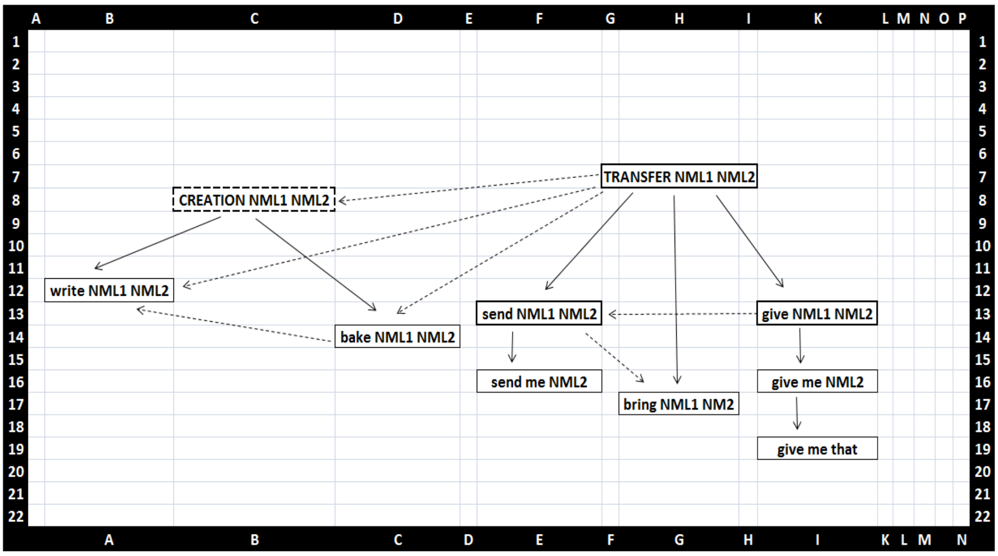


Figure 2.17: the ditransitive complex category; 4 of 5.

37. CREATION-NML1-NML2 ⟶ Rob built Mary a house.

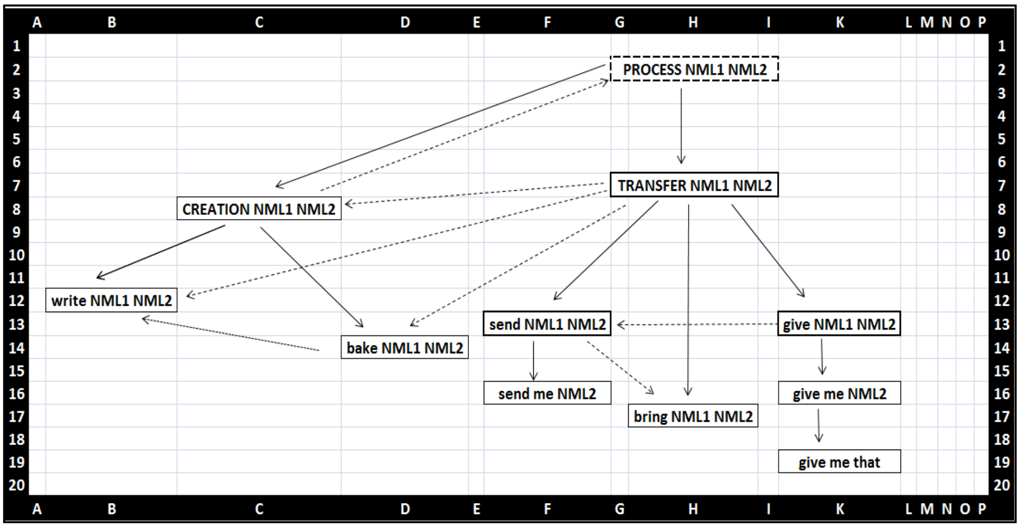


Figure 2.18: the ditransitive complex category; 5 of 5.

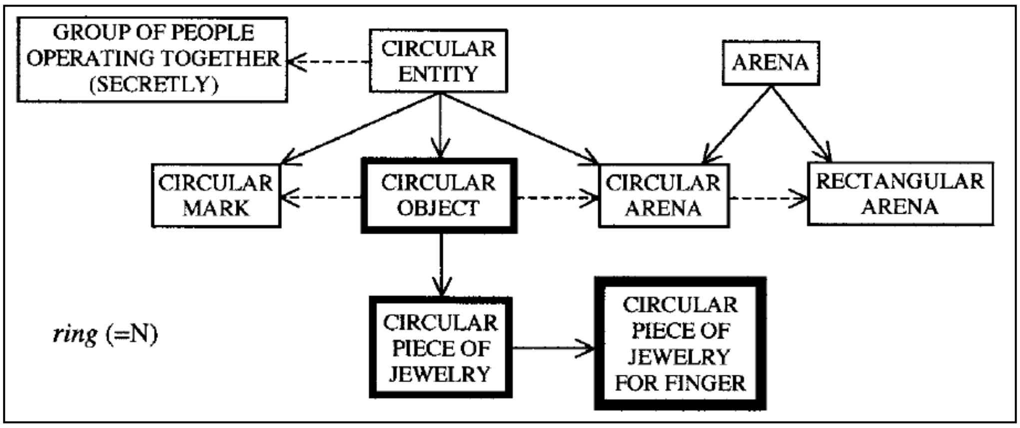


Figure 2.19: the polysemous lexical item *ring* represented as a complex category (from Langacker, 2008, 37, fig. 2.2). Dashed arrows indicate relationships of extension. Solid arrows indicate relationships of elaboration/instantiation. Thickness of boxes indicates the degree of entrenchment of the units enclosed in them.

19.1.3. A USAGE-BASED APPROACH TO LANGUAGE ACQUISITION

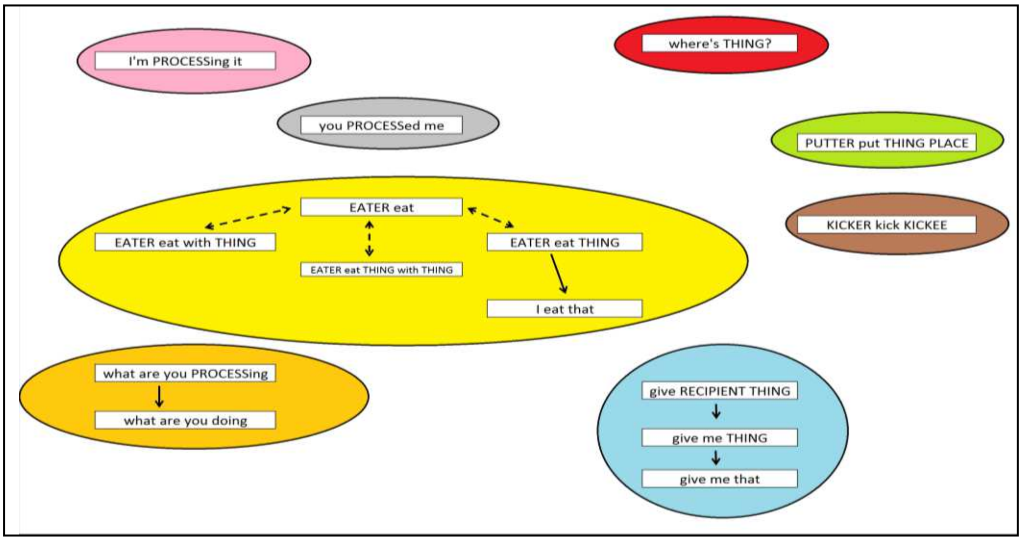


Figure 3.1: Children's inventory of constructions. Each construction can be thought of as a mini-grammar representing lexically-specific knowledge. Symbolic units are enclosed in rectangles in which *small letters* indicate elements that are concrete (phonologically specified) and *CAPITAL LETTERS* indicate slots. Dashed arrows indicate relationships of extension, whereas solid arrows indicate relationships of elaboration.

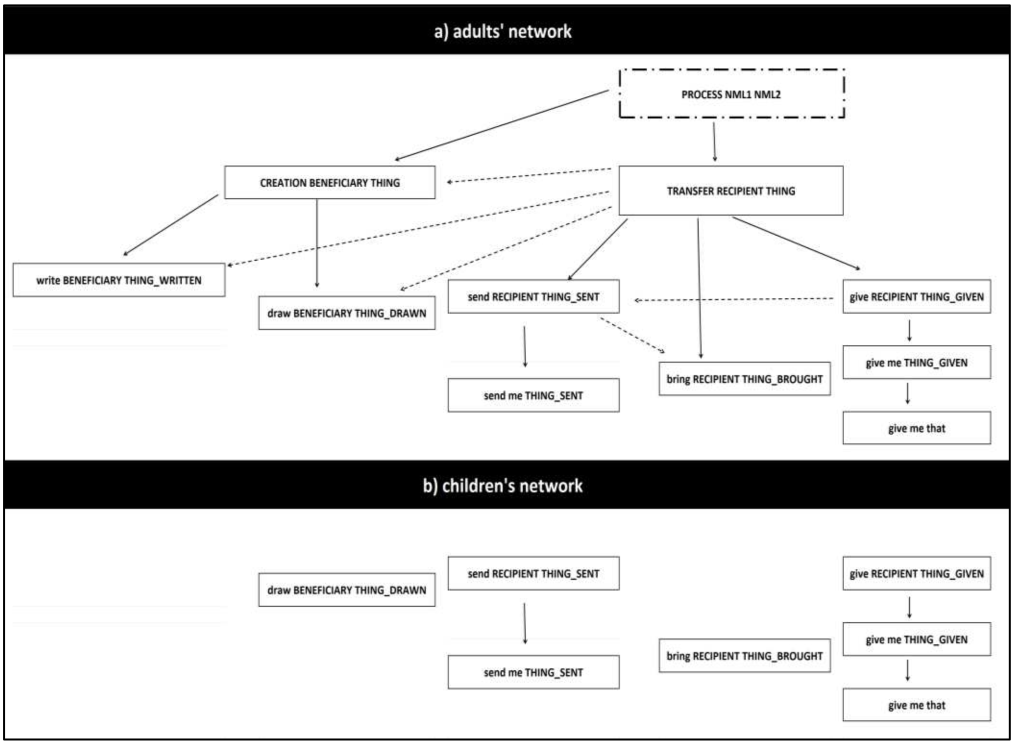


Figure 3.2: Adults' (a) and children's (b) representation of the ditransitive construction. Children lack the more schematic layers and their units are fewer and much more poorly interconnected. Dashed arrows indicate relationships of extension, whereas solid arrows indicate relationships of elaboration. Symbolic units are enclosed in rectangles. *Small letters* indicate lexically-specific elements, whereas *CAPITAL LETTERS* indicate slots. Thickness of rectangles indicates the degree of entrenchment of the units contained in them.

1. Now I want the big box of sweets.

2. *a.* ***I want*** *my book.*

   b. ***I want*** *ice-cream.*

3. I'll get you ***the big box of sweets***.

4. Mum is tired, ***now***.

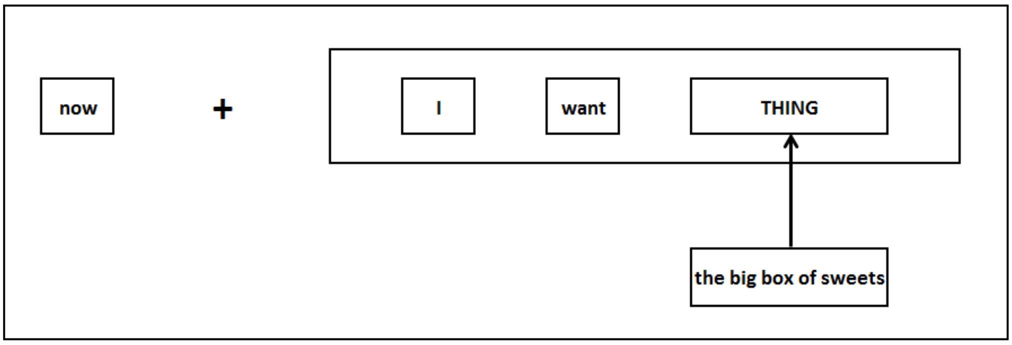


Figure 3.3: the traceback method: deriving *Now I want the big box of sweets* by superimposing (arrow) *I want THING* and *the big box of sweets* and by juxtaposing *now* (+).

5. I want I eat an apple.

6. 1 *CHI: grandma, I'll gorp you.
   2 *GRD: you what?
   3 *MOT: I <gorp> [!] you.
   4 *MOT: it is a kind of greeting that goes on in a cartoon
   5 Rob watches a lot.
   6 *CHI: 0 [=! gorps his grandmother and laughs].
   7 *MOT: grandma, what happened to you?
   8 *GRM: I got gorped by Rob.

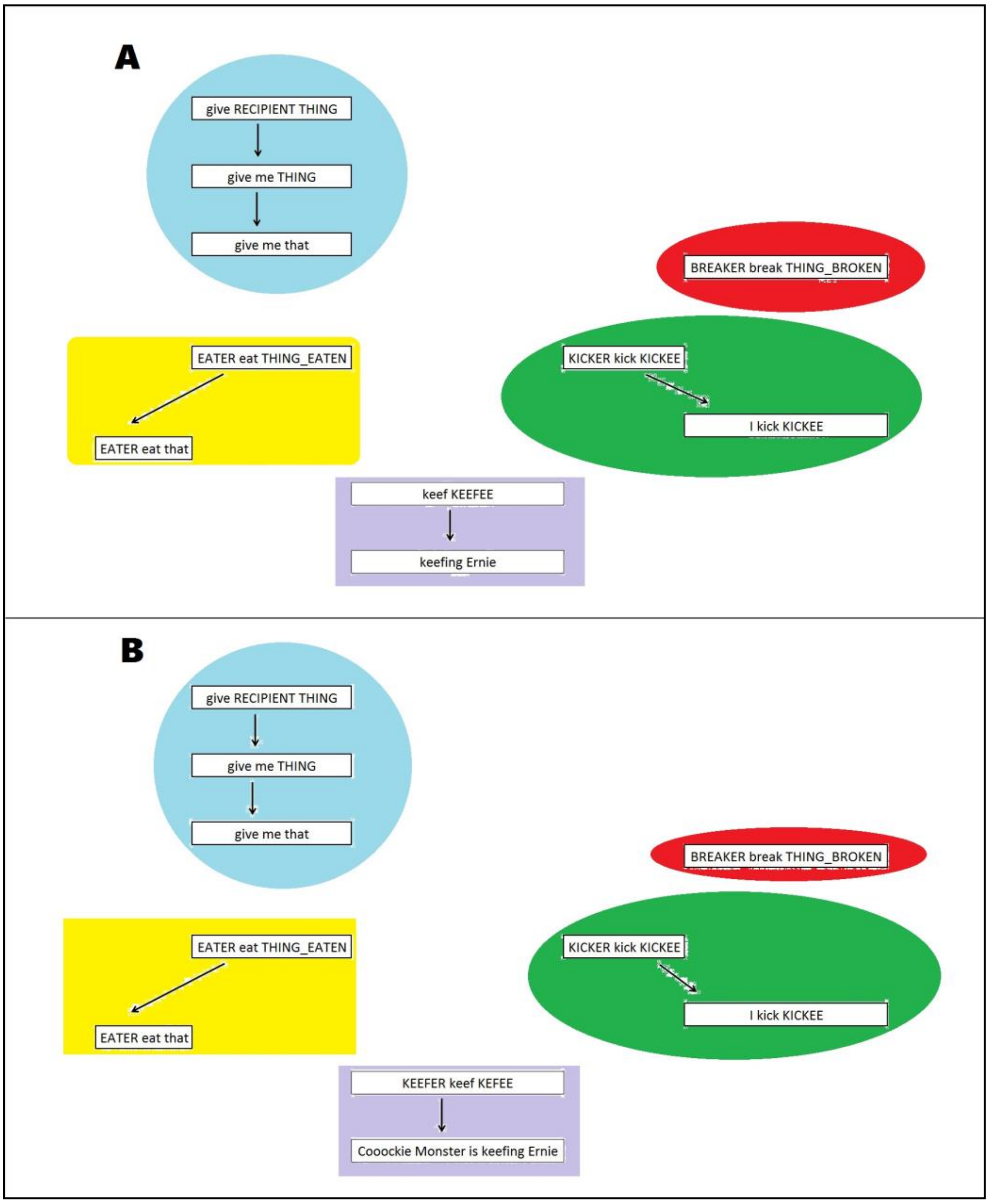


Figure 3.4: the verb *keef* enters children's lexically-specific inventories under VO (a) and SVO (b) conditions. Symbolic units are enclosed in rectangles in which *small letters* indicate lexically-specific material and *CAPITAL LETTERS* indicate slots. Arrows indicate relationships of elaboration/instantiation.

7. SVO

   Elmo dacking the car

   Elmo dacked the car

8. SOV

   Elmo the car gopping

   Elmo the car gopped

9. VSO

Tamming Elmo the car

Tamed Elmo the car

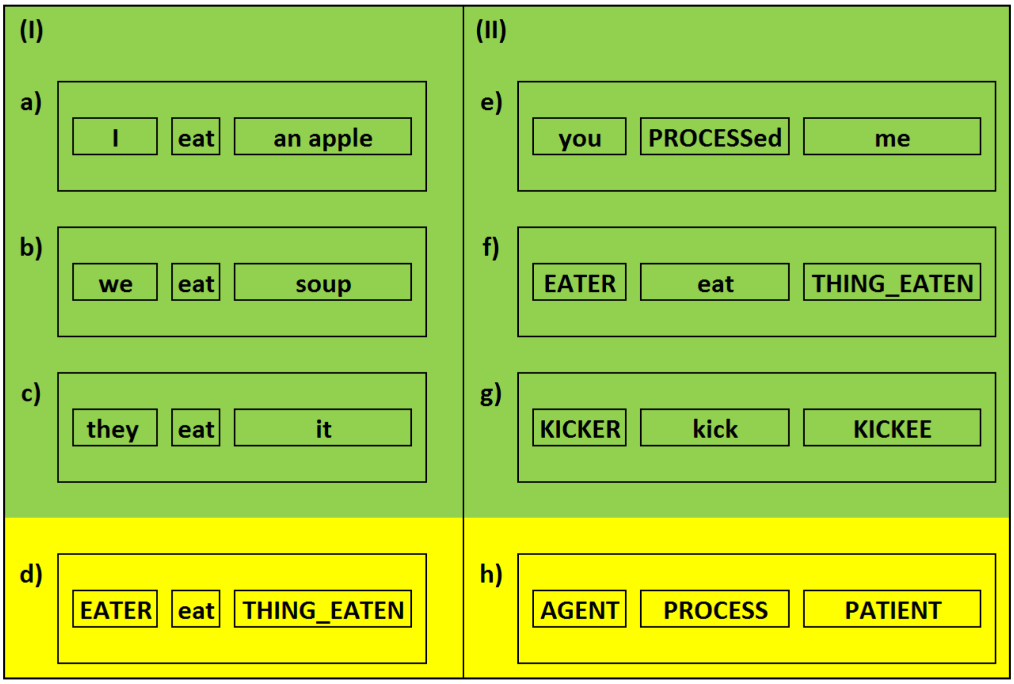


Figure 3.5 the development of the constructional schema *AGENT-PROCESS-PATIENT* (II), which is inferred by applying the same processeses of functional distributional analysis and structural alignment that underline the development of lexically-specific schemas (I).

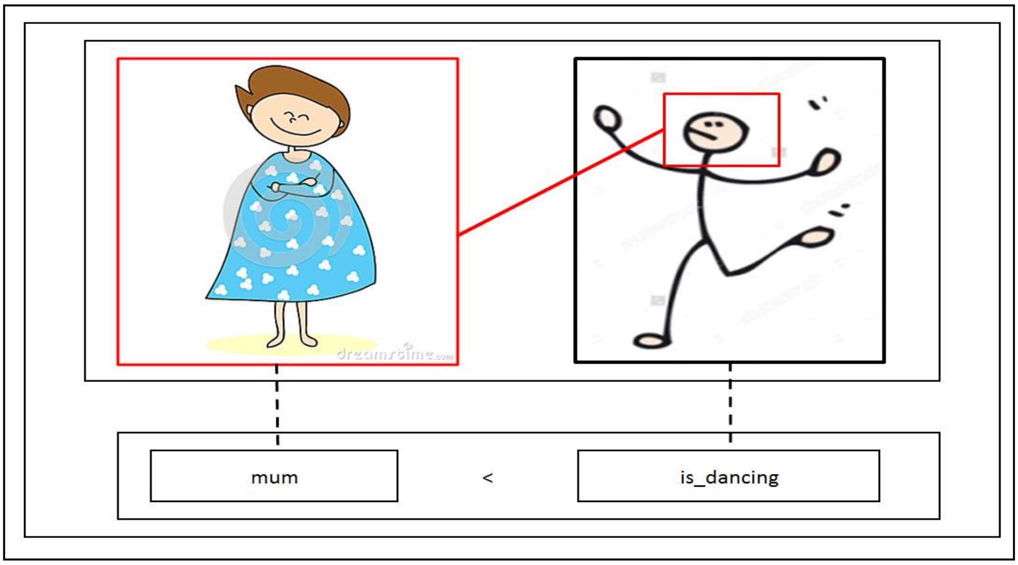


Figure 3.6: the symbolic unit *mum is dancing*. The phonological pole (bottom) is reported in standard spelling. The upper row, where the drawings are, represents the semantic pole. Dashed lines represent relationships of symbolisation. The red box surrounding the sketched man's head indicates that the meaning of *dancing* implies a dancing entity. The fact that the semantic pole of /mum/ is in red and is linked by a red line to the man's head (enclosed in a red box) indicates that *mum* is the dancing entity. The symbol "<" indicates linear order at the phonological pole.

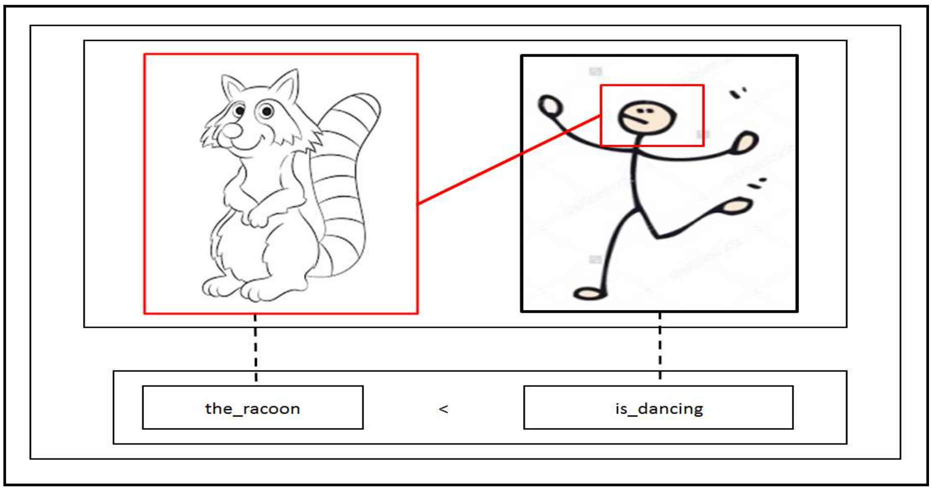


Figure 3.7: the symbolic unit *the racoon is dancing*. See the previous figure on how to interpret this figure.

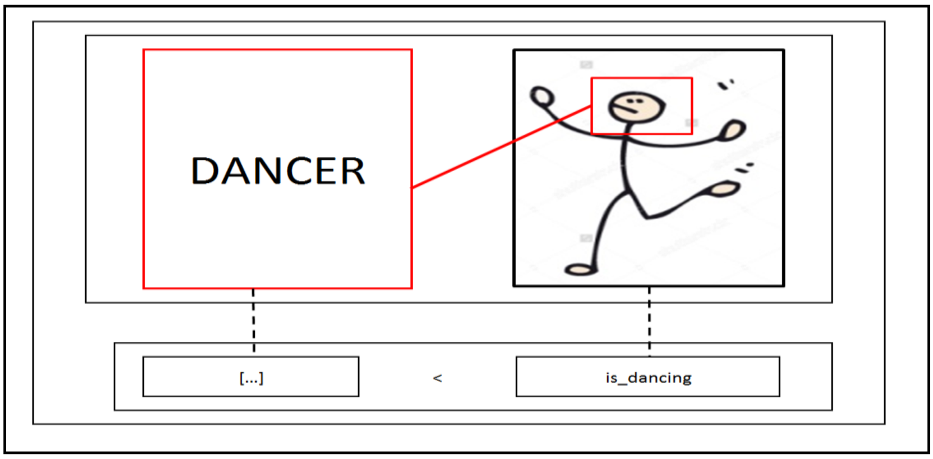


Figure 3.8: The schema *DANCER-is-dancing*. The interpretation of the figure is identical to fig. 3.6 and 3.7. What changes is that the dancing entity is specified only schematically at the semantic pole (*DANCER*) and not specified at all at the phonological pole ([…]).

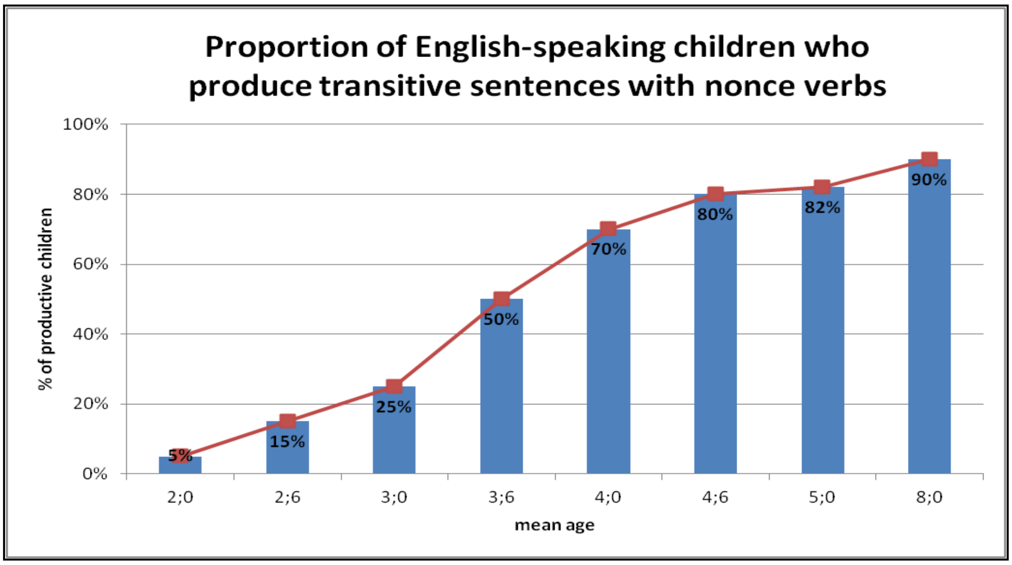


Figure 3.9: Proportion of English-speaking children who produce adultlike transitive sentences with nonce verbs. Based on Tomasello (2006b, fig. 6.1, p. 266).

10. a) mum put your pacifier on the table.

b) I moved the chair into the living room.

19.1.4. DESIGN RATIONALE

a) *Can Italian-speaking children's early language be accounted for in terms of lexically-specific units acquired from the concrete language that children themselves have previously experienced?*

b) *To what extent can Italian-speaking children be said to rely on (have mastered) fully-schematic constructions/patterns?*

1. *Mangialo!*
   mangi-a=l-o!
   eat-IMP.2.SG=clitic.3.ACC-M.SG
   "Eat it/him!"

2. *L' ha mangiato*

l'	*ha*	*mangi-a-t-o*
clitic.3.ACC	has	eat-TV(conj.I)-PTCP-M.SG

   "(She/he) ate it/him."

19.2. Part II: THE SPONTANEOUS PRODUCTION OF A TWO-YEAR-OLD CHILD

19.2.1. METHOD

Table 6.1: more qualitative measures of the MacArthur Questionnaire. Roberto's results are compared with both his peers' and thirty-month-olds' average results.

LANGUAGE WITHOUT EXTRA-LINGUISTIC CONTEXT	Roberto	26-27-month-olds who often do so	30-month-olds who often do so
is the child able to understand language without extra-linguistic context? (about objects and people not present, future and past events)	often	70%	90%
is the child able to speak about things without extra-linguistic context? (about objects and people not present, future and past events)	often	50%	more than 70%
MORPHOLOGICAL COMPLEXITY	**Roberto**	**proportion of 26-27-month-olds who show morphological complexity**	**proportion of 30-month-olds who show morphological complexity**
does the child use both singular and plural forms of the same noun?	yes	70%	87.5%
does the child use various forms of the same adjective (number and gender inflections)?	yes	55%	81.3%
does the child use various forms of the verb (mood-tense-person markers)?	yes	67.5%	87.3%
GRAMMATICAL COMPLEXITY	**Roberto**	**26-27-month-olds**	**30-month-olds**
use of functional words, adverbial phrases, co-ordination and subordination (*daddy went away with the car*) as opposed to paratactic strategies (daddy away car) - % of complex sentences	100%	less than 50%	less than 80%
use of pronouns (*I want this*) as opposed to nouns (*Roberto wants this*) - % of pronoun usages	66%	more than 30%	more than 50%

1. a) **CHI:* *<l' ho trovato> [?].*
l' *ho* *trov-a-t-o*
clitic.3.ACC have(PRS.1.SG) find-TV(conj.I)-PTCP-M.SG
"I found it."
(week6.2014.02.18.B: line 393)

b) **CHI:* *<non xxx> [=! whispering].*
non xxx
not xxx
"Don't xxx."
(week6.2014.02.18.B: line 2729)

2. **CHI: stivaletti so(no) +//.*
Stival-ett-i *so(no) +//.*
Boot-little(DIM)-M.PL be(PRS.3.PL)
"Little boots are +//."
(week.6.2014.02.18.B: line 1224)

3. **CHI: voio [: voglio] [* p] &*RES:sì attaccare (.) <un> [//] questo e voio [: voglio] [* p] +//.*
voglio attaccare questo e voglio +//.
vogli-o attacc-a-re quest-o
want-PRS.1.SG attach-TV(conj.I)-INF this-M.SG
e vogli-o +//.
and want-PRS.1SG
"I want to attach this and I want +//."
(week6.2014.02.18.B: lines 1334-1335)

Table 6.2: Roberto's test corpus production: identifying multi-word sentences.

sentence type	no.	%
intelligible multi-word sentences	993	52%
one-word sentences	401	21%
non intelligible multi-word sentences	390	21%
interrupted sentences	106	6%
total	**1890**	**100%**

4. a) *MOT: *un'altra?*
un' altr-a ?
a other-F.SG ?
"Another one?"

b) *CHI: *un'altra*
un' altr-a
a other-F.SG
"Another one."

(week.6.2014.02.18.A.: lines 84 and 85).

5. a) *CHI: *col motorino.*
co=l motorin-o
with=the(M.SG) scooter-M.SG
"With the scooter."

b) *CHI: *col motorino.*
co=l motorin-o
with=the(M.SG) scooter-M.SG
"With the scooter."

(week6. 204.02.18.B: lines 76 and 81)

6. a) *RES: *no*, *questa* ***è una salciccia.***
no, *quest-a* ***è*** ***un-a*** ***salsicci-a***
no, this-F.SG **is** **a-F.SG** **sausage-F.SG**
"No, this **is a sausage**."

b) *CHI: ***è una salciccia.***
è ***un-a*** ***salsicci-a***
is ***a-F.SG*** ***sausage-F.SG***
"(It) **is a sausage**."

(week6.2014.02.18.B: lines 1692 and 1693)

7. a) *CHI: ***lo porta via*** *il chiamoncino [: camioncino] [* p].*
l-o ***port-a*** ***via*** *il*
clitic.3.ACC-M.SG **take-PRS.3.SG** **away** the(M.SG)
camion-cin-o
truck-little(DIM)-M.SG
"The little truck, (he) **takes it away**."

b) *CHI: ***lo porta via.***
l-o ***port-a*** ***via***
clitic.3.ACC-M.SG **take-PRS.3.SG** **away**
"(He) **takes it away**."

(week6.2014.18.A: lines 1369 and 1370)

8. a) *RES: *ah@i, ha scavato e ha tirato sù un tubo.*
ha *scav-a-to* *e* *ha* *tir-a-to*
has dig-TV(conj.I)-PTCP and has pick-TV(conj.I)-PTCP
sù *un* *tub-o*
up a(M.SG) pipe-M.SG
"(He) dug and picked (pulled) up a pipe."

b) *CHI: *ha scavato e ha tirato (.) sù un tubo e ha fatto tcsh@o.*
ha *scav-a-to* *e* *ha* *tir-a-to*
has dig-TV(conj.I)-PTCP and has pick-TV(conj.I)-PTCP
sù *un* *tub-o* *e* *ha* *fatto* *tcsh*
up a(M.SG) piper-M.SG and has done *tcsh*
"(He) dug and picked (pulled) up a piper and went 'tcsh'."

(week6.2014.02.18.B: lines 3098 and 3099)

9. a) *RES: *<l' ho> [/] l' ho visto uno morto.*
l' *ho* *vis-t-o*
clitic.3.ACC have(PRS.1.SG) see-PTCP-M.SG
un-o *mort-o*
one-M.SG dead-M.SG
"I've seen a dead one."

b) *CHI: *l' hai visto uno morto.*
l' *hai* *vis-t-o*
clitic.3.ACC have(PRS.2.SG) see-PTCP-M.SG
un-o *mort-o*
one-M.SG dead-M.SG
"You've seen a dead one."

(week6.2014.02.18.B: lines 2616 and 2617)

10. a) *ADL: *il papà ieri è andato a lavorare*
il *papà* *ieri* *è* *and-a-t-o*
the(M.SG) daddy(M.) yesterday is go-TV(conj.I)-PTCP-M.SG
a *lavor-a-re*
to work- TV(conj.I)-INF
"Yesterday Daddy went to work."

b) *CHI: *il papà è andato a lavorare*
il *papà* *è* *and-a-t-o* *a*
the(M.SG) daddy(M.) is go-TV(conj.I)-PTCP-M.SG to
lavor-a-re
work- TV(conj.I)-INF
"Daddy went to work."

11. a) *MOT: *e poi prende l' aereoplano e* ***va <lontano*** *> [/] lontano „ vero?*
e *poi* *prend-e* *l'* *areoplan-o* *e*
and then take-PRS.3.SG the airplane-M.SG and
va ***lontano*** *lontano,* *vero?*
goes **far** far , true?
"And then (he) takes the airplane and goes far away, right?"

b) *CHI: **va lontano**
va ***lontano.***
goes **far.**
"(He) goes away."

(week6.2015.02.18.A: lines 542 and 551)

12. a) *CHI: *il lupo mangia tutto.*

il	*lup-o*	*mangi-a*	*tutt-o*
the(M.SG)	wolf-M.SG	eat-PRS.3.SG	all-M.SG

b) *CHI: *mangia <tutto> [!] il lupo.*

mangi-a	*tutt-o*	*il*	*lup-o*
eat-PRS.3.SG	all-MSG	the(M.SG)	wolf-M.SG

"The wolf eats everything."

(week6.2014.02.18.B: lines 2825 and 2831)

Table 6.3: Roberto's intelligible multi-word-sentences: identifying target sentences.

Roberto's intelligible multi-word sentences	no.	%
novel multi-word sentences (target sentences)	768	77%
instantiations of an already considered target sentence type	78	8%
songs, poems and nursery rhymes	70	7%
partial and full imitations	52	5%
partial and full self-repeats	25	3%
TOTAL	**993**	**100%**

13. TARGET SENTENCE

*CHI: *no, qui, **così non si fa!***

no, qui,	***così***	***non***	***si***	***fa***
no, here	**like.that**	**not**	**clitic.IMPRS.NOM**	**does**

"No, here, that's not the way to do it."

(week6.2014.02.18.B: line 3088)

14. UNIT AND ITS INSTANTIATIONS

a) ***Così*** + ***non<si<fa***

così	***non***	***si***	***fa***
like.that	**not**	**clitic.IMPRS.NOM**	**does**

"That's not the way to do it."

b) *FAT: *digli (..) **così non si fa .***

d-i=gli	***così***	***non***
tell-IMP.2.SG=clitic.3.DAT	**like.that**	**not**

si	***fa***
clitic.IMPRS.NOM	**does**

"Tell him, that's not the way to do it."

(week1.2014.01.12.B.chat: line 180)

c) *CHI: ***non si fa così !***

non	***si***	***fa***	***così***
not	**clitic.IMPRS.NOM**	**does**	**like.that**

"That's not the way to do it."

(week3.2014.01.21: line 2081)

15. TARGET SENTENCE
*CHI: *e* ***lì va*** *que [: questo] [* p].*
e ***lì*** ***va*** *quest-o*
and **there** **goes** this-M.SG
"And this one goes there."
(week6.2014.02.18.B: line 2363)

16. PUTATIVE UNITS
a) *FAT: *<va lì> [<]* !
va lì
goes there
"(It) goes there."
(week6.2014.02.12: line 1197)

b) *MOT: *eh@i , e va lì .*
eh, e va lì
eh, and goes there
"Eh, and (it) goes there."
(week3.2014.01.21: line 2698)

17. a) FULLY-SPECIFIC PACKET
non<si<fa + così
non si fa così
not clitic.IMPRS.NOM does like_that
"That's not the way to do it."

b) FIXED STRING
va lì
va lì
goes there
"(It) goes there."

18. *CHI: *tra poco <il lupo [: Luca] [* s:ur] > [//]* ***il Luca vas [: va] [* p] <ca [: casa] [* p -ret]> [//] a casa***
Tra poco ***il Luca va a casa***
tra poco ***il Luca va a cas-a***
between little **the(M.SG) Luca goes to home-F.SG**
("Luca goes home in a little while").
"Luca will go home in a while/ Luca is going home soon."
(week6.2014.18.B: line 2889-2890)

19. a) *MOT: *il Luca va a casa !*

il	***Luca***	***va***	***a***	***cas-a***
the(M.SG)	**Luca**	**goes**	**to**	**home-F.SG**

"Luca goes home."
(week5.2014.02.04: line 2006)

b) *MOT: ***il Luca*** *adesso* ***va a casa***

il	***Luca***	*adesso*	***va***	***a***	***cas-a***
the(M.SG)	**Luca**	now	**goes**	**to**	**home-F.SG**

"Now, Luca is going home."
(week2.2014.01.14.A: line 193)

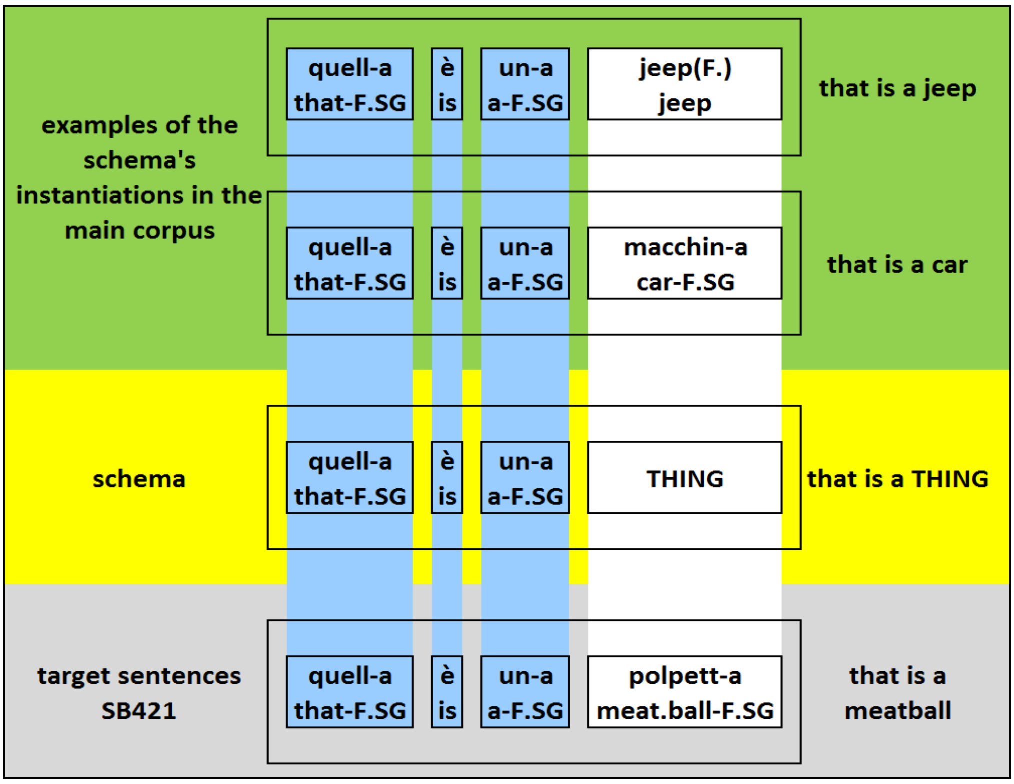


Figure 6.1: target sentence (grey strip), schema (yellow strip) and the schema's instantiations (green strips). The slots and its instantiations are highlighted in white, whereas the lexically specific part of the schema is highlighted in blue.

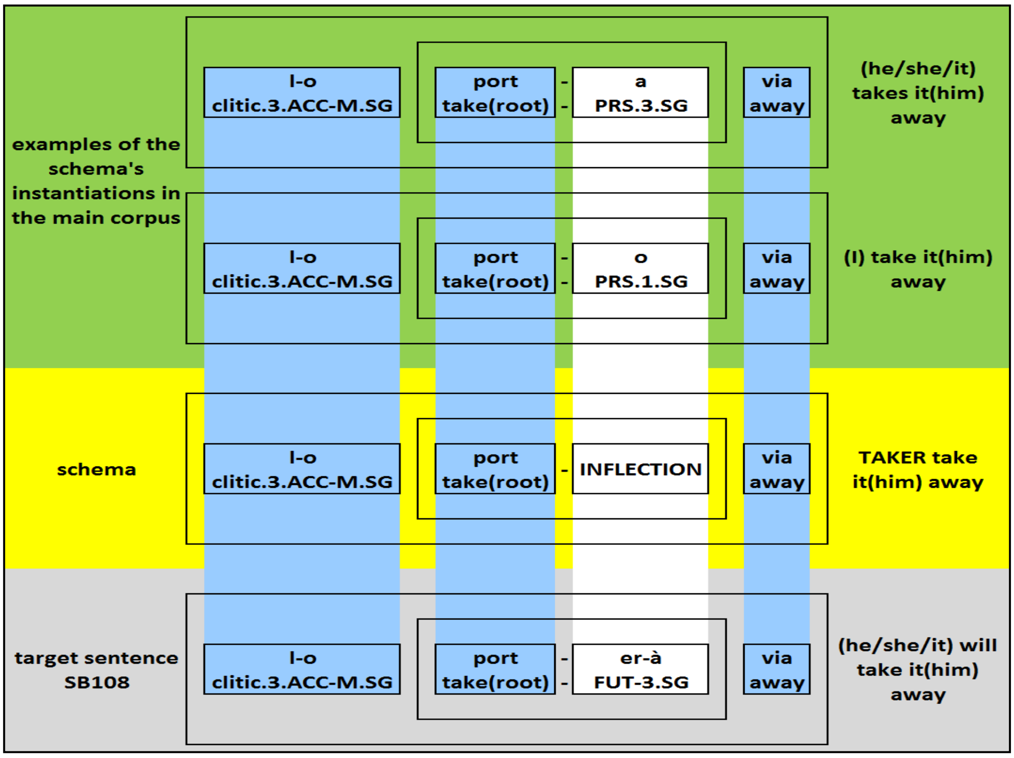


Figure 6.2: see fig. 6.1 above on how to read this figure.

20. FIXED-SCHEMAS

a) *Questa è una THING*
quest-a è un-a THING
this-F.SG is a-F.SG THING
"This is a THING."

b) *Lo port-INFLECTION via*
l-o port-INFLECTION away
clitic.3.ACC-M.SG take(root)-INFLECTION away
"TAKER take it away."

21. SCHEMATIC-PACKETS

a) La<THING + è<caduta [S+V]
l-a THING è cad-u-t-a
the-F.SG THING is fall-TV(conj.II)-PTPC-F.SG
"The<THING + has<fallen."

*b)*CHI: è caduta la tenda* [VS]
è cad-u-t-a l-a tend-a
is fall-TV(conj.II)-PTPC-F.SG the-F.SG curtain-F.SG
"The curtain has fallen."
(week2.2014.01.16: line 889)

*c)*RES: la torre è caduta* [SV]
l-a torr-e è cad-u-t-a
the-F.SG tower(F.)-SG is fall-TV(conj.II)-PTPC-F.SG
"The tower has fallen."
(week3.2014.01.21: line 1889)

22. a) *Il<treno + sta<per<partire*
il tren-o st-a per
the(M.SG) train-M.SG stay-PRS.3.SG to
part-i-re
leave-TV(conj.III)-INF
“The train is about to leave.”

23. a) *FAT: *il treno sta per partire*
il tren-o st-a per
the(M.SG) train-M.SG stay-PRS.3.SG to
part-i-re
leave-TV(conj.III)-INF

b) *FAT: *sta per partire il treno*
st-a per part-i-re il
stay-PRS.3.SG to leave-TV(conj.III)-INF the(M.SG)
tren-o
train-M.SG

“The train is about to leave.”

(week5.2014.02.07: lines 2541 and 2542)

24. *CHI: *voglio questo*
vogli-o quest-o
want-PRS.1.SG this-M.SG
“I want this one.”
(fictitious example)

25. a) *CHI: *il pane voglio*
il pan-e vogli-o
the(M.SG) bread(M.)-SG want-PRS.1.SG
“I want bread.”
(fictitious, line 1)

b) *CHI: *voglio il pane*
vogli-o il pan-e
want-PRS.1.SG the(M.SG) bread(M.)-SG
“I want bread.”
(fictitious, line 5)

26. Voglio + THING_WANTED
vogli-o + THING_WANTED
want-PRS.1.SG + THING_WANTED
“I want THING_WANTED.”

27. **CHI: quella voglio*
quell-a vogli-o
that-F.SG want-PRS.1.SG
"I want that one."
(fictitious example)

28. *THING_WANTED voglio*
THING_WANTED vogli-o
THING_WANTED want-PRS.1.SG
"I want THING_WANTED."

29. a) *CHI: ***stava <lavorando (..)> [>] lassù.***
st-a-v-a *lavor-a-ndo*
stay-TV(conj.I)-IMPERF-3.SG work-TV(conj.I)-ing
la+(s)sù
there+up
"(He/she/it) was working up there."
(week6.2014.02.18.A: line 1192)

b) *Stava lavorando*
st-a-v-a *lavor-a-ndo*
stay-TV(conj.I)-IMPERF-3.SG work-TV(conj.I)-ing
"(He/she/it) was working."

c) ***Lassù.***
la+(s)sù
there+up
"Up there."

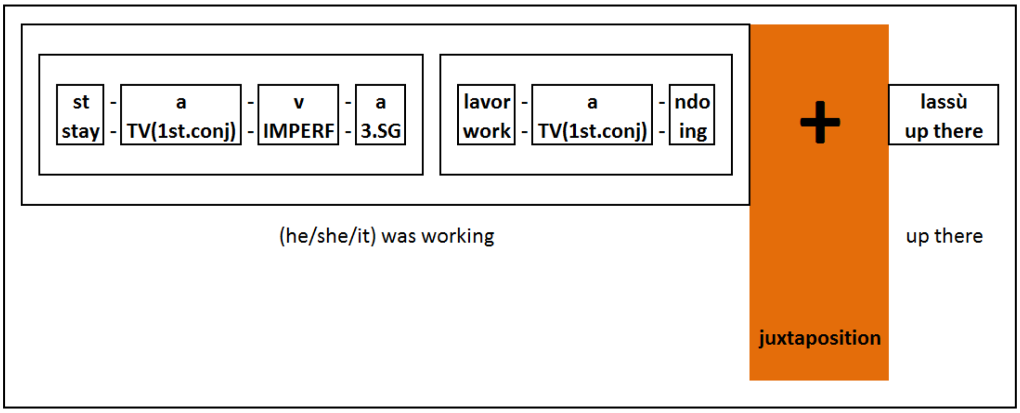


Figure 6.3: the juxtaposition of unit_A and unit_B. The juxtaposition of the two units is coded with a plus (+) sign and is highlighted in red.

30. **CHI: ho preso questi*
ho *preso* *quest-i*
have(PRS.1.SG) taken this-M.PL
"I took these."
(week6.2014.02.18.A: line 756)

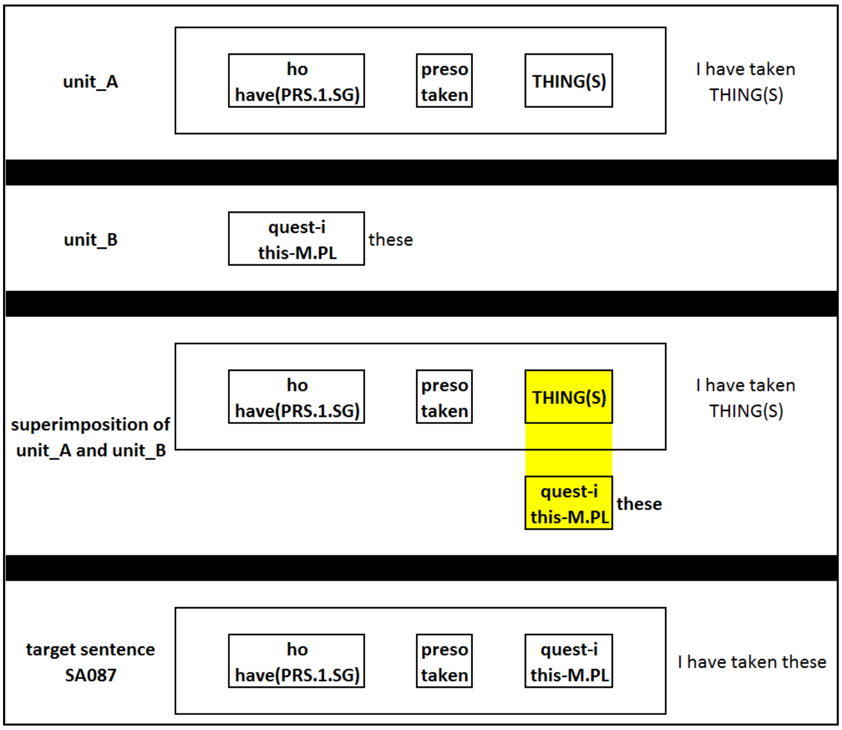


Figure 6.4: illustrating the operation of superimposition. Slot elaboration is highlighted in yellow.

31. TARGET SENTENCE (SB108)
    *CHI: *lo porterà via.*
    l-o *port-er-à* *via*
    clitic.3.ACC-M.SG take-FUT-3.SG away
    "(S/he/it) will take it/him away."
    (week6.2014.02.18.B: line 897)

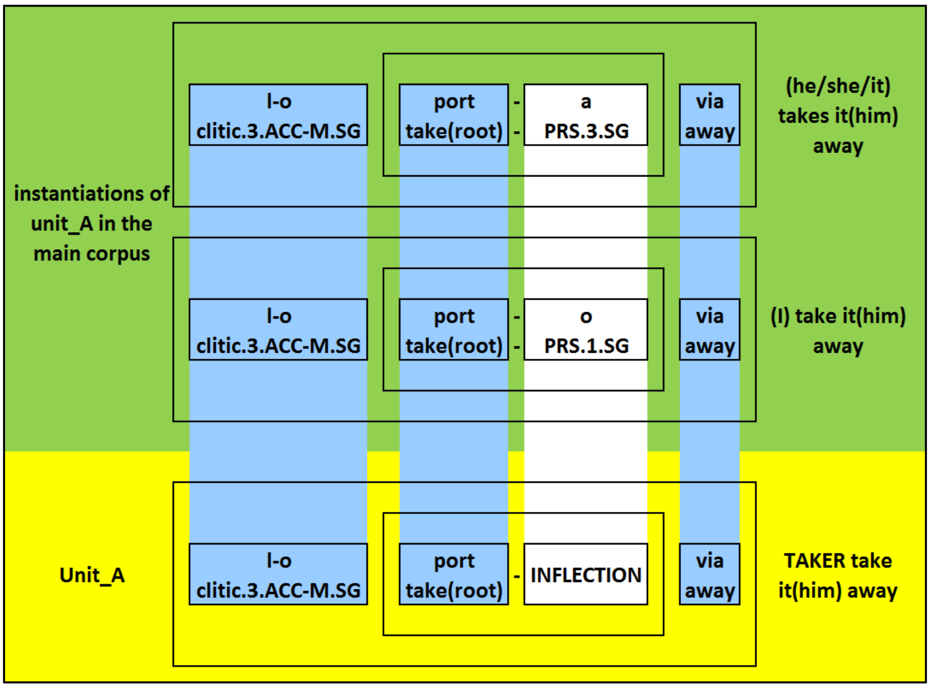


Figure 6.5: Component unit_A of target sentence SB108 (31) (yellow strip) and its instantiations in the main corpus (green strips). Slot formation is highlighted in white. Shared and fixed lexical material is highlighted in blue.

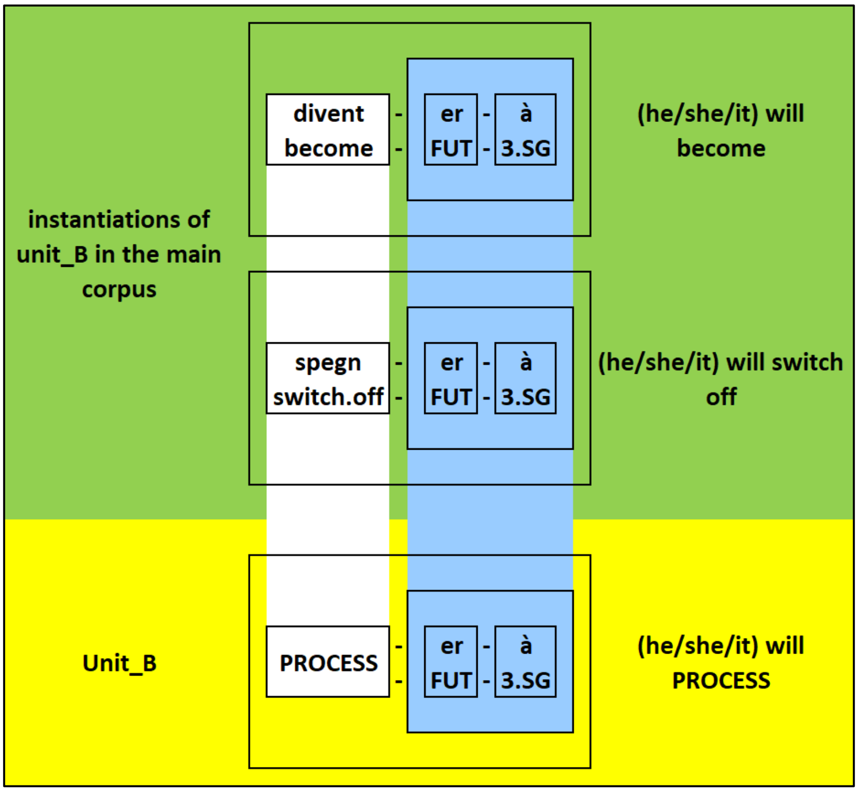


Figure 6.6: Component unit_B of target sentence SB108 (31) (yellow strip) and its instantiations in the main corpus (green strips). See previous figures on how to interpret how colours are used.

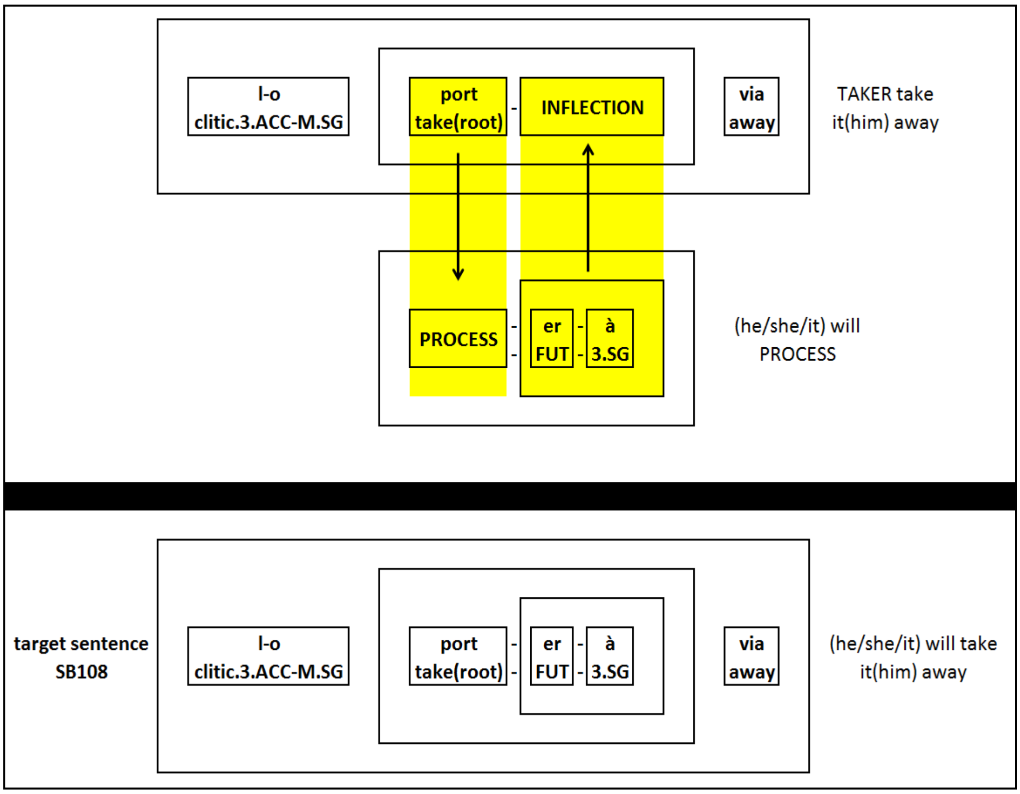


Figure 6.7: deriving target sentence SB108 (31) through mutual superimposition of the schema fig. 6.7 (unit_B) and the schema fig. 6.6 (unit_A). Arrows move from the filler to the elaborated slot.

32. TARGET SENTENCE (SA014)

*CHI: ***cosa c' è dentro qua?***

cosa c(i) è dentro qua?

what there.clitic.LOC is inside here

"What's inside here?"

(week6.2014.02.18.A: line 243)

33. COMPONENT UNITS

a) *CHI: ***cosa c' è dentro qua ?***

cosa c(i) è dentro qua?

what there.clitic.LOC is inside here

"What's inside here?"

(week6.2014.02.17: line 161).

b)*CHI: *sai* ***coa [: cosa] [* p] c' è dentro qua ?***

sa-i ***cosa c(i) è dentro qua?***

know-PRS.2.SG **what there.clit.LOC is inside here**

"Do you know what's inside here?"

(week6.2014.02.17: line 161)

*34. *CHI: l'ho lasciato a casa della nonna*

l' ho lasci-a-t-o

clitic.3.ACC have(PRS.1.SG) leave-TV(conj.I)-PTCP-M.SG

a cas-a de=ll-a nonn-a

to home-F.SG of=the-F.SG grandma-F.SG

"I've left it at grandma's."

(week6.2014.02.14.A: line 1175).

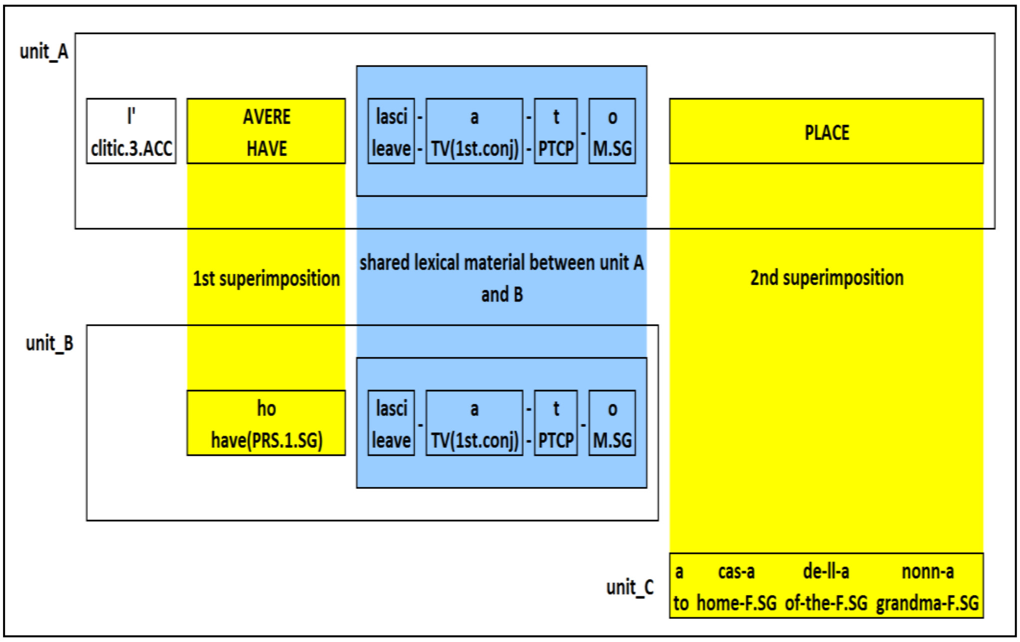


Figure 6.8: deriving target sentence SA146 (34) through two superimpositions (in yellow). The lexically-specific material shared by the two units is highlighted in blue.

35. Mum is dancing.

36. a) La mamma sta ballando

l-a	*mamm-a*	*st-a*	*ball-a-ndo*
the-F.SG	mum-F.SG	stay-PRS.3.SG	dance-TV(conj.I)-ing

b) *Sta ballando, la mamma*

st-a	*ball-a-ndo*	*l-a*	*mamm-a*
stay-PRS.3.SG	dance-TV(conj.I)-ing	the-F.SG	mum-F.SG

“Mum is dancing.”

37. a) *MOT: *lo diamo a Roberto*

l-o	*d-iamo*	*a*	*Roberto*
clitic.3.ACC-M.SG	give-PRS.1PL	to	Roberto

“We give it to Roberto.”

b) *CHI: *a Roberto lo diamo*

a	*Roberto*	*l-o*	*d-iamo*
to	Roberto	clitic.3.ACC-M.SG	give-PRS.1.PL

“We give it to Roberto.”

(fictitious examples)

38. a) *Il maialino*

il	*maial-in-o*
the(M.SG)	pig-little(DIM)-M.SG

b) *Il piccolo maiale*

il	*piccol-o*	*maial-e*
the(M.SG)	little-M.SG	pig(M.)-SG

c) *Il maiale piccolo*

il	*maial-e*	*piccol-o*
the(M.SG)	pig(M.)-SG	little-M.SG

“The little pig.”

39. the little THING.

40. Il THING-in-o

il	*THING-in-o*
the(M.SG)	THING-little(DIM)-M.SG

“The little THING.”

41. I eat an apple.

42. I eat pasta.

43. I-eat-THING_EATEN.

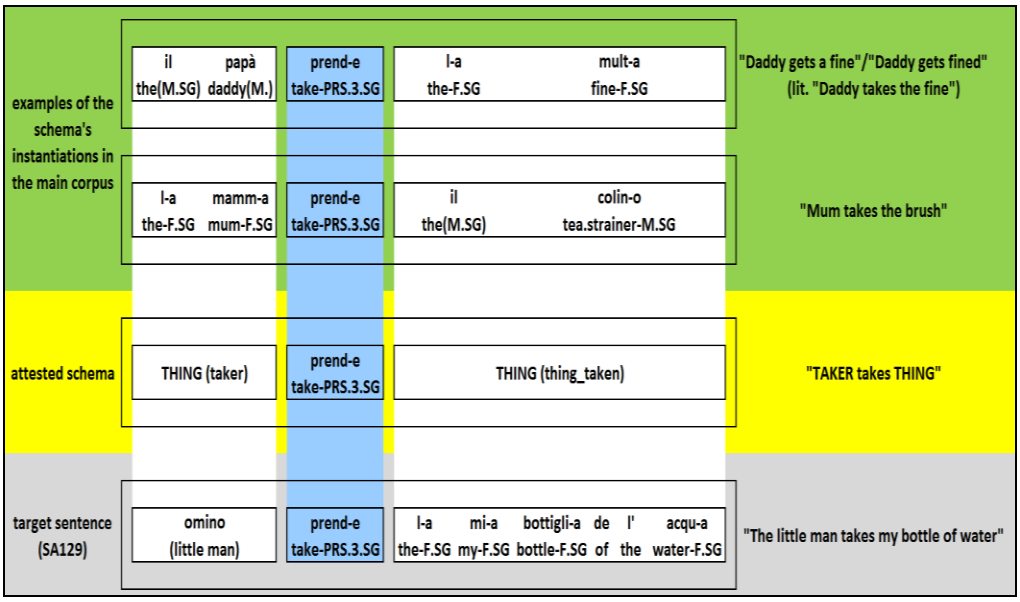


Figure 6.9: a schema (yellow) with two *THING* slots: *THING prende THING_TAKEN* "THING takes THING_TAKEN". Slot formation is in white. Recurring lexical material is in blue

44. *Mamma prend-e i pomodor-i*

mamm-a	*prend-e*	*i*	*pomodor-i*
mum-F.SG	take-PRS.3.SG	the(M.PL)	tomato-M.PL

"Mum takes the tomatoes."

45. **Loro prende questo*

loro	*prend-e*	*quest-o*
they(NOM.3.PL)	take-PRS.3.SG	this-M.SG

"*They takes this."

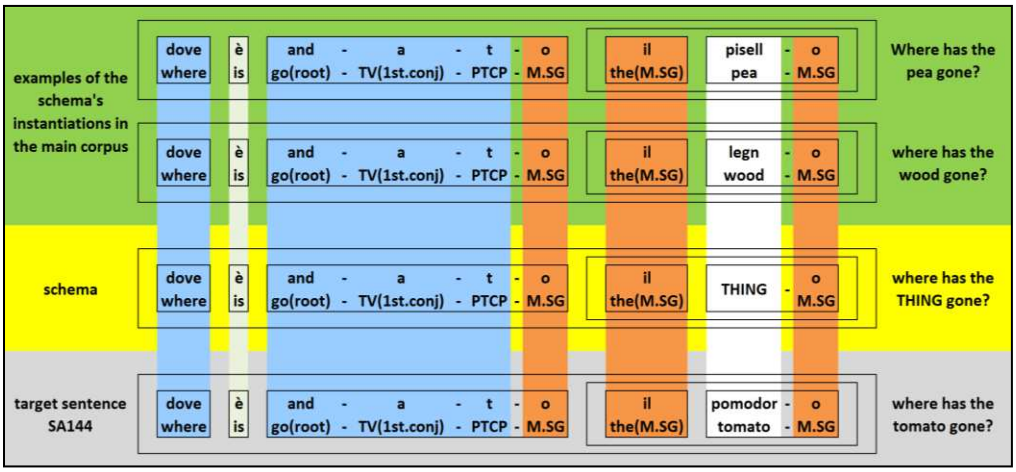


Figure 6.10: Explaining gender and number agreement: the parts highlighted in red show the morphologically-specified, recurring elements in schema (yellow strip), instantiations (green strip) and target sentences (dark grey strip) that account for both gender and number agreement. The light grey highlighting indicates fixed, recurring elements that account for number agreement (namely the PRS.3.SG of the verb *to be*). White parts indicate the slot and its instantiations. Blue parts indicate other shared lexical material, not relevant for agreement.

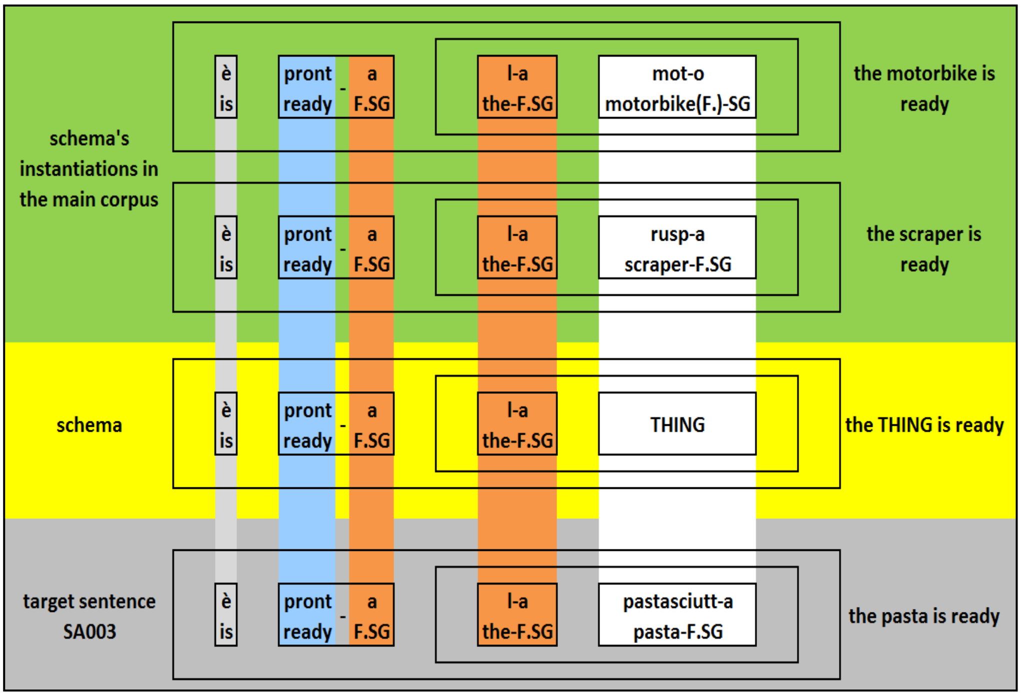


Figure 6.11: see fig. 6.10 on how to read this figure.

46. * È *pronta la risotto*
    è *pront-**a*** *l-**a*** *risott-**o***
    is ready-**F.SG** the-**F.SG** rice-**M.-SG**
    "The rice is ready."

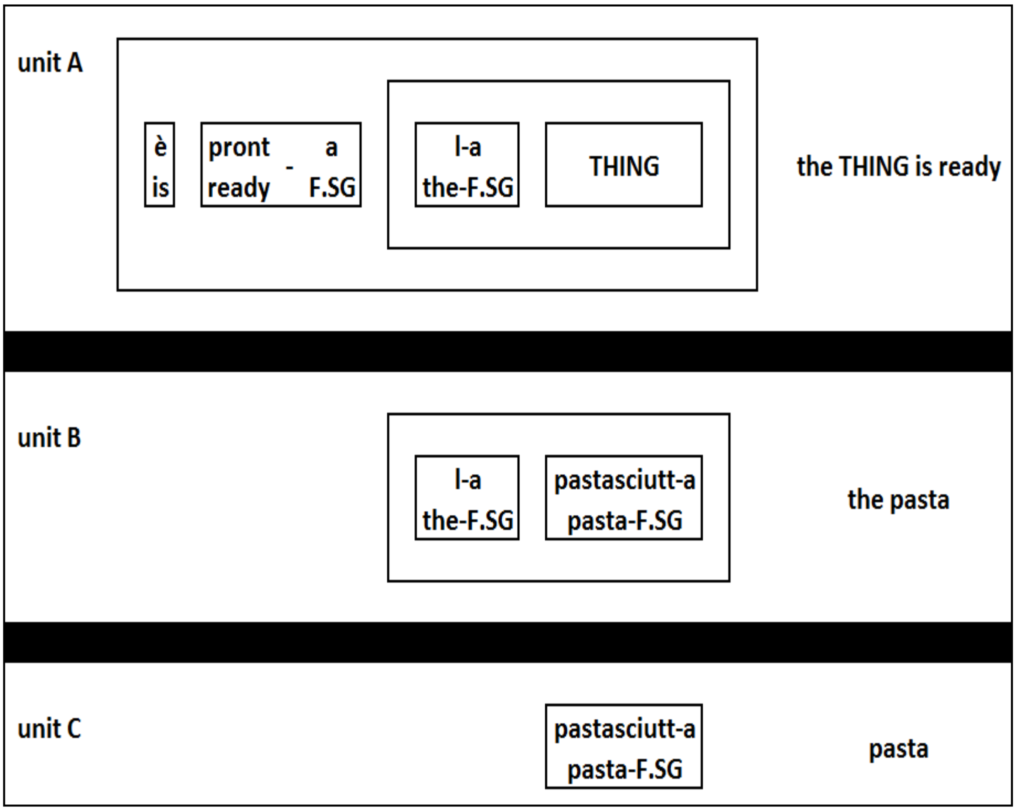


Figure 6.12: Three possible component units of target sentence SA003.

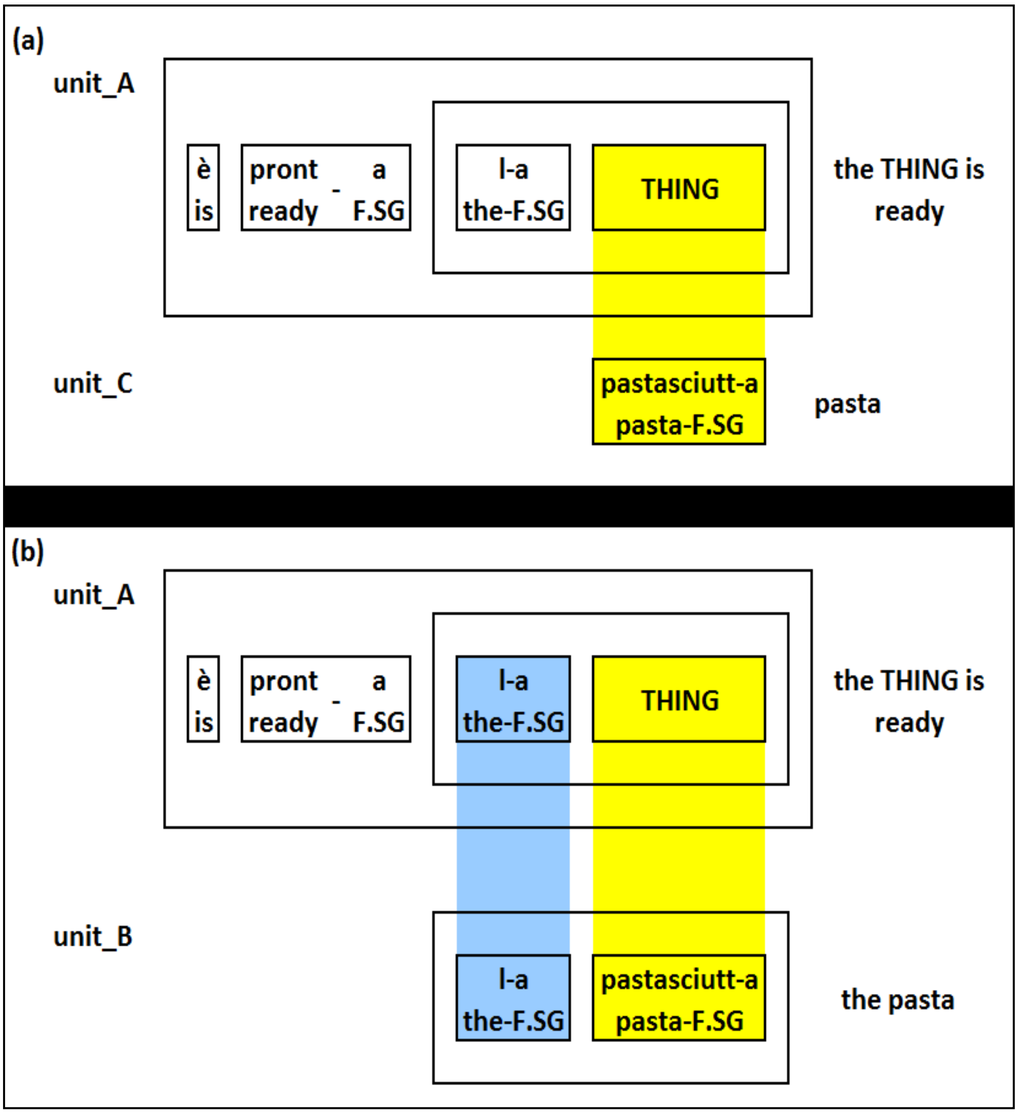


Figure 6.13: Two ways of deriving target sentence SA003 in fig. 6.12. Slot elaboration is highlighted in yellow. Shared lexical material is highlighted in blue.

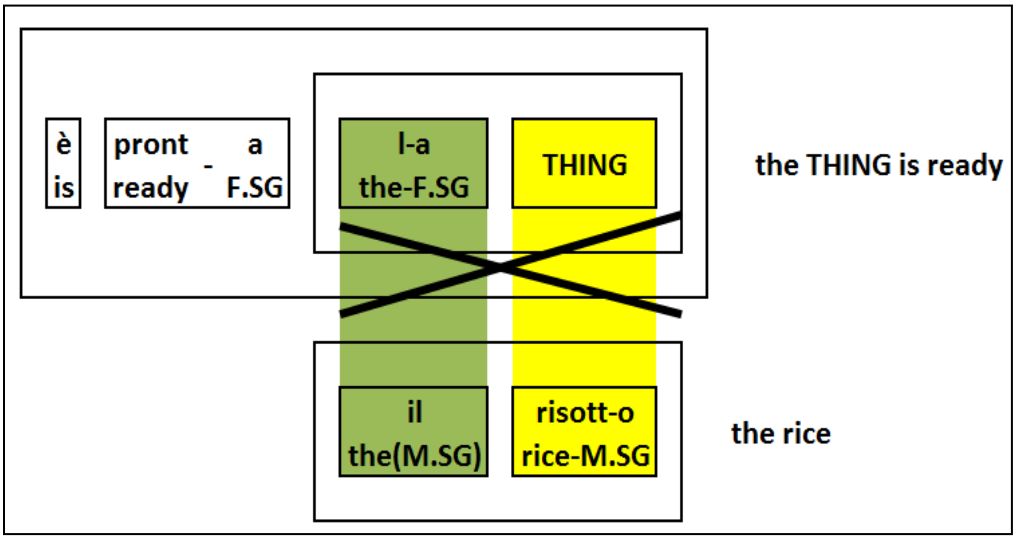


Figure 6.14: activating larger units prevents ungrammatical sentences.

19.2.2. RESULTS

Table 7.1: Results: using different traceback methods to analyse Roberto's target sentences. Method_A is the method adopted for the main analysis (refer back to chapter 6). For each method, the frequency threshold a precedent had to meet in order to be considered as an available component unit and who could have uttered it are reported in the "method's description". Hence, in Method_C, target sentences were traced back to strings that the child (and the child only) uttered in the main corpus at least once (excluding imitations and self-repeats).

METHOD	method's description	number of operations needed for successful derivations					fails	total no. of utterances
		exact match (0 operation)	1	2	3	4 or more		
Method_A	2 any speaker	113	219	142	90	69	135	768
Method_B	3 any speaker	94	205	135	101	71	162	768
Method_C	1 child only	109	216	131	91	55	166	768
Method_D	1 any speaker	144	252	140	99	59	74	768

METHOD	method's description	type of operation needed		
		sup	juxt	tot
Method_A	2 any speaker	1062	25	1087
Method_B	3 any speaker	1091	23	1114
Method_C	1 child only	984	18	1002
Method_D	1 any speaker	1069	19	1088

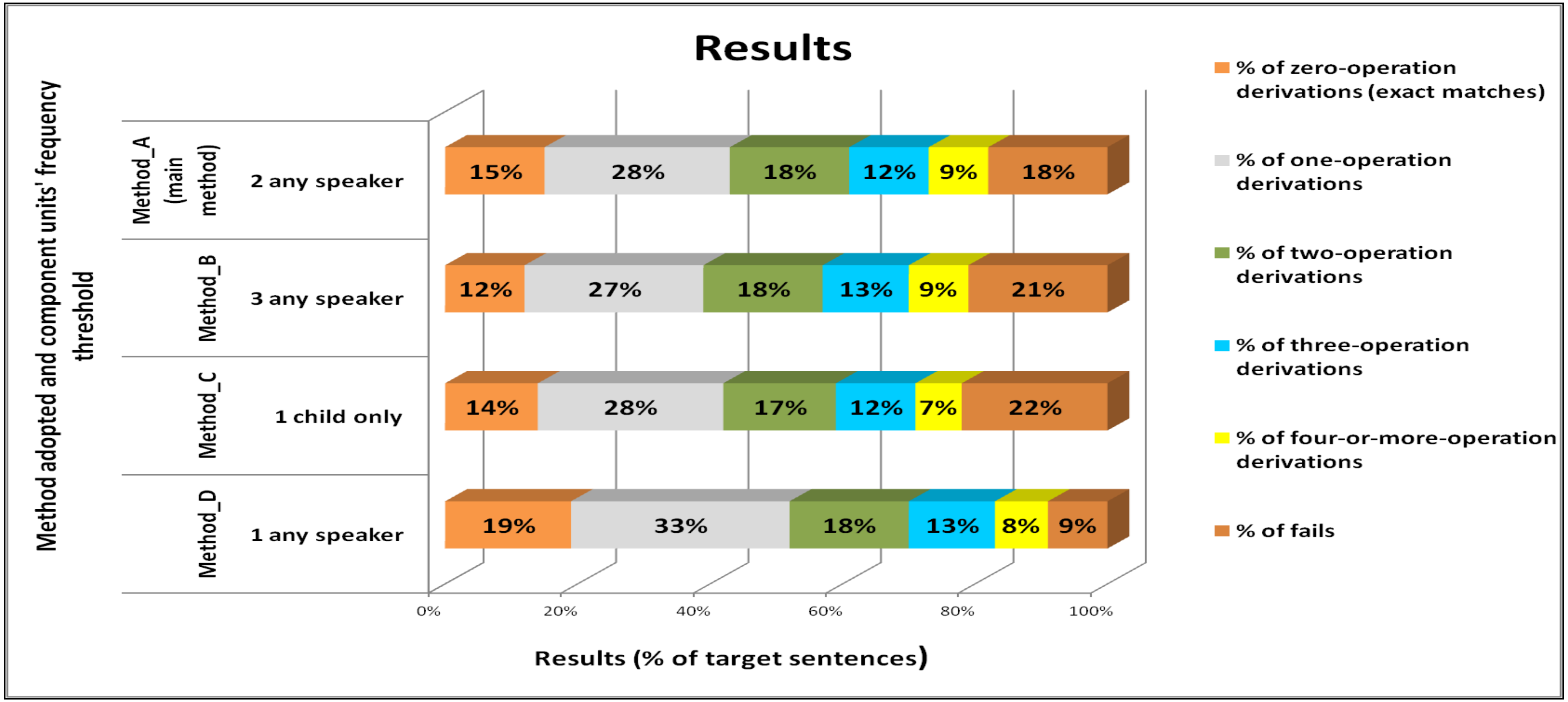


Figure 7.1: Comparing the results of different traceback methods. Method_A is the method adopted for the main analysis (refer back to chapter 6). For each method, the frequency threshold a precedent had to meet in order to be considered as an available component unit and who could have uttered it are reported on the left hand-side of each bar. Hence, in Method_C, target sentences were traced back to strings that the child (and the child only) uttered in the main corpus at least once (excluding imitations and self-repeats).

Table 7.2: results as a proportion of all (993) Roberto's intelligible multi-word sentences using Method_A.

TYPE OF SENTENCE	no.	%	is the sentence novel/creative?
two-and-three-operation derivations	232	23%	yes
one-operation derivations	219	22%	yes
fails	135	14%	yes
zero-operation derivations (exact matches)	113	11%	no
instantiations of an already attested target sentence type	78	8%	no
imitations and self-repeats (full and partial)	77	8%	no
songs, poems and nursey rhymes	70	7%	no
four-or-more operation derivations	69	7%	yes
total	**993**	**100%**	

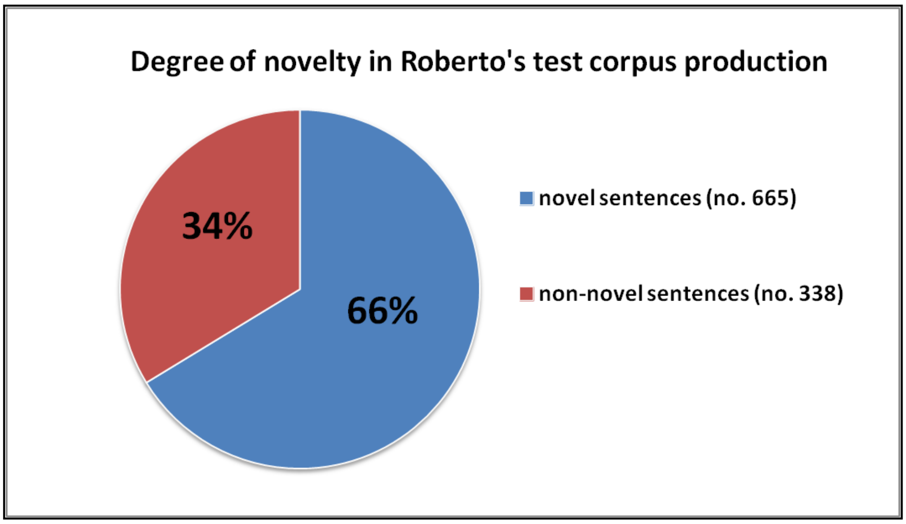


Figure 7.2: degree of novelty in Roberto's test corpus intelligible multi-word sentences.

47. *CHI: <sai> [/] sai che io (.) ho fatto una corsa (..) grande + e sono arrivato (.) dal Luca +. +. e ho fatto pf@o.*
sai che io ho fatto una corsa grande e sono arrivato dal Luca e ho fatto puf.
sa-i *che*
know-PRS.2.SG that
io *ho* *fatto* *un-a* *cors-a* *grand-e*
I have(PRS.1.SG) done a-F.SG run-F.SG big-SG
e *sono* *arriv-a-t-o* *da=l*
and be(PRS.1.SG) arrive-TV(conj.I)-PTCP-M.SG at=the(M.SG)
Luca *e* *ho* *fatto* *puff.*
Luca and have(PRS.1.SG) done puff.
("You know that I have done a big run and I have arrived at the Luca and I've done 'puff'?").
"Did you know that I ran a long way and got to Luca and went 'puff'?"
(week5.2014.02.18.B: lines 1423 and 1425)

Table 7.3: Target Sentences that required four or more operations to be derived from their component units.

sentences that required four or more operations to be derived	no.	% of 69 sentences that required four or more operations	% of 768 total target sentences
four-operation derivations	45	65%	6%
five-operation derivations	17	25%	2%
six-operation derivations	4	6%	0.5%
seven-operation derivations	1		
eight-operation derivations	1	4%	0.5%
ten-operation derivations	1		
total	**69**	**100%**	**9%**

Table 7.4: Distribution of fail types.

Method_A: fail type	no. of target sentences that could not be sucessfully derived	% of 768 target sentences	% of 135 fails
constructional	90	12%	67%
lexical	38	5%	28%
both constructional and lexical	7	1%	5%
total	**135**	**18%**	**100%**

48. *CHI: *è (..) una cera.*
è un-a cer-a
is a-F.SG wax-FG
"(It) is a wax."
(week.6.2014.02.18.A: line 304)

49. *È una THING-a*
è un-a THING-a
is a-FSG THING-F.SG
"(It) is a THING."

Table 7.5: Constructional fails of Method_A analysed with other methods.

Constructional fails of Method_A analysed with other traceback methods					
Method	method's description	Successful	**lexical	*constructional	total (no. of sentences)
Method_A	2 any speaker	0	0	97	97
Method_C	1 child only	27	7	63	97
Method_D	1 any speaker	44	3	50	97

*sentences that contain a constructional fail or contain a constructional and a lexical fail

**sentences that do not classify as constructional fails but that contain a lexical fail

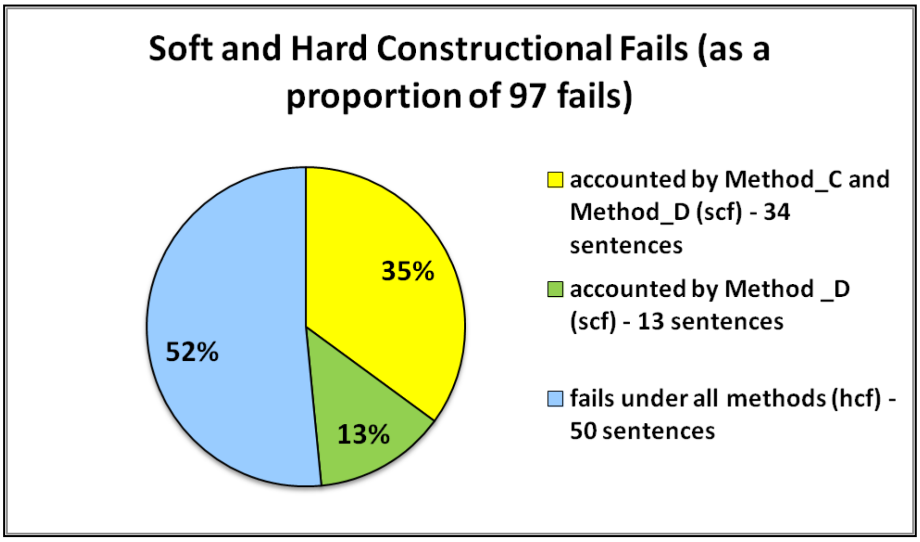


Figure 7.3: Soft (scf) and Hard (hcf) Constructional Fails.

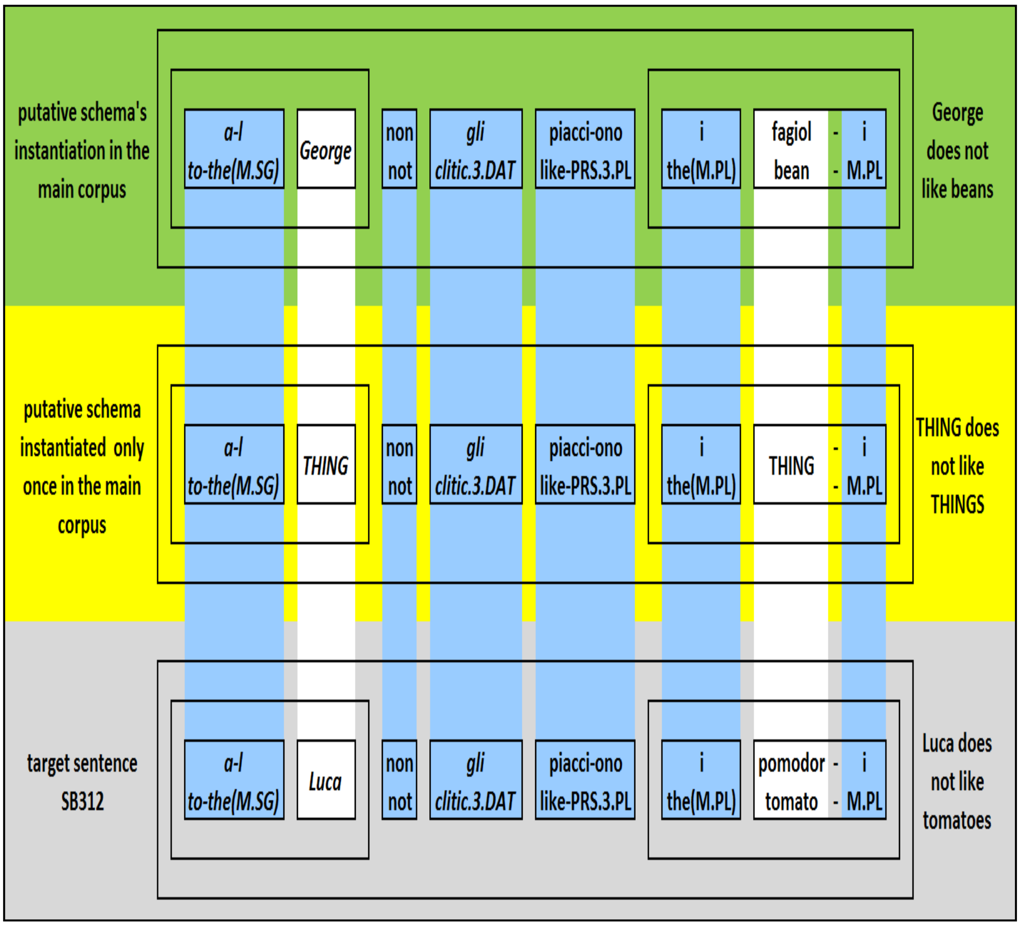


Figure 7.4: Target sentence SB312 (grey strip), its only putative precedent (green strip) and the schema they both instantiate (yellow strip). Slots are in white, whereas shared concrete material is in blue. Elements in *italics* are co-indexed. For the type of construction used in target sentence SB312, refer back to 1.4.3, sentences (42)-(45).

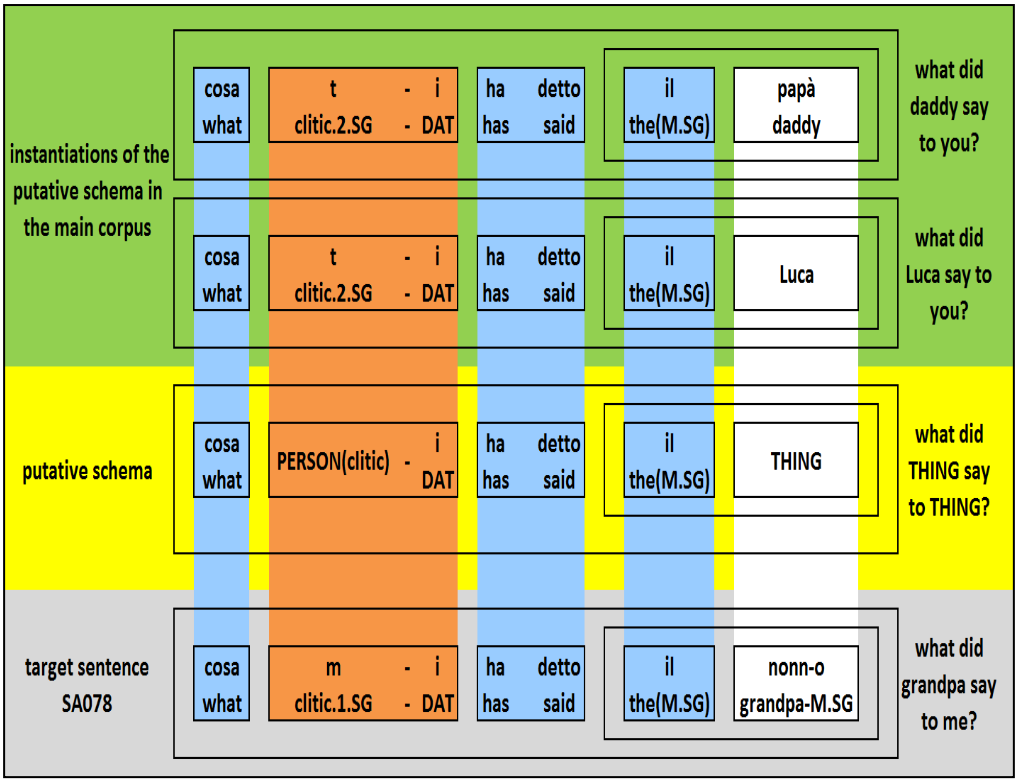


Figure 7.5: Constructional fails at the clause level. Target sentence SA078 (grey strip), its precedents (green strip) and the putative schema they instantiate (yellow strip). Highlighted in red is the putative slot of the schema which does not meet the type variance requirements to be considered as such. Successful slot formation is highlighted in white and recurring lexical material is in blue.

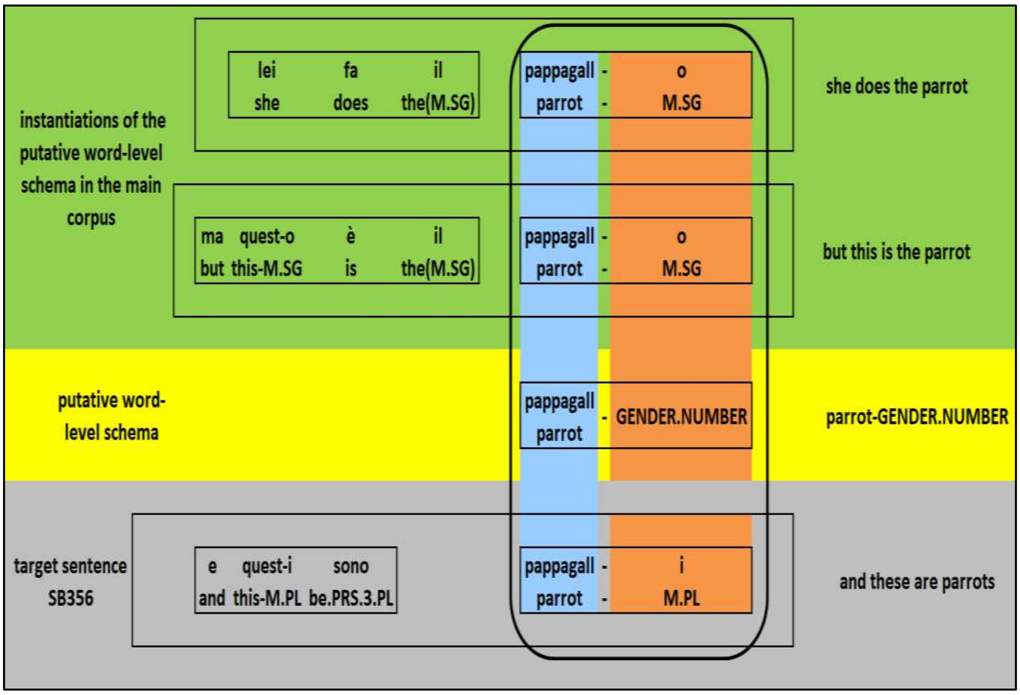


Figure 7.6: Constructional fails at the word level (i.e. MORPHOLOGICAL FAILS). Target sentence SB356 (grey strip). The schema that does not meet the type variance requirement (yellow strip), its instantiations (green strip) and the relevant part of the target sentence are enclosed in the rounded rectangle. Highlighted in red is the putative slot of the schema which does not meet the type variance requirements to be considered as such. The fixed part of the schema is highlighted in blue.

Table 7.6: Syntactic, morphological and lexical fails under Method_A.

fail type	no. of sentences containing at least one fail	% of 135 fails	% of 768 target sentences
syntactic fails	77	57%	10%
lexical fails	38	28%	5%
morphological fails	11	8%	1.5%
both syntactic and lexical fails	5	4%	1%
both syntactic and morphological fails	2	2%	0.1%
both morphological and lexical fails	2	1%	0.1%
total	**135**	**100%**	**17.7%**

Table 7.7: Types of Fully Lexically-specific Strings.

Types of Fully Lexically-specific Strings	no.	%
fixed-strings whose internal order is the only possible grammatical sequence	186	37%
single words	156	31%
fixed-strings	147	29%
onomatopoeias	7	2%
packets	6	1%
total	**502**	**100%**

50. *C' era una volta una bella favolina.*
c(i)[1] *er[2]-a[3]* *un[4]-a[5]*
there(clitic.LOC) [1] be(IMPERF)[2]-3.SG[3] a[4]-F.SG[5]
volt[6]-a[7] *un[8]-a[9]* *bell[10]-a[11]*
time[6]-F.SG[7] a[8]-F.SG[9] nice[10]-F.SG[11]
favol[12]<in[13]>-a[14]
fairy.tale[12]<little(DIM)[13]>-F.SG[14]
"Once upon a time there was a little fairy tale."

Table 7.8: types of schema with slots.

Types of Schemas-with-Slot	no.	%
multi-word fixed-schemas	634	91%
single-word schemas	35	5%
multi-word schematic-packets	29	4%
total	698	100%

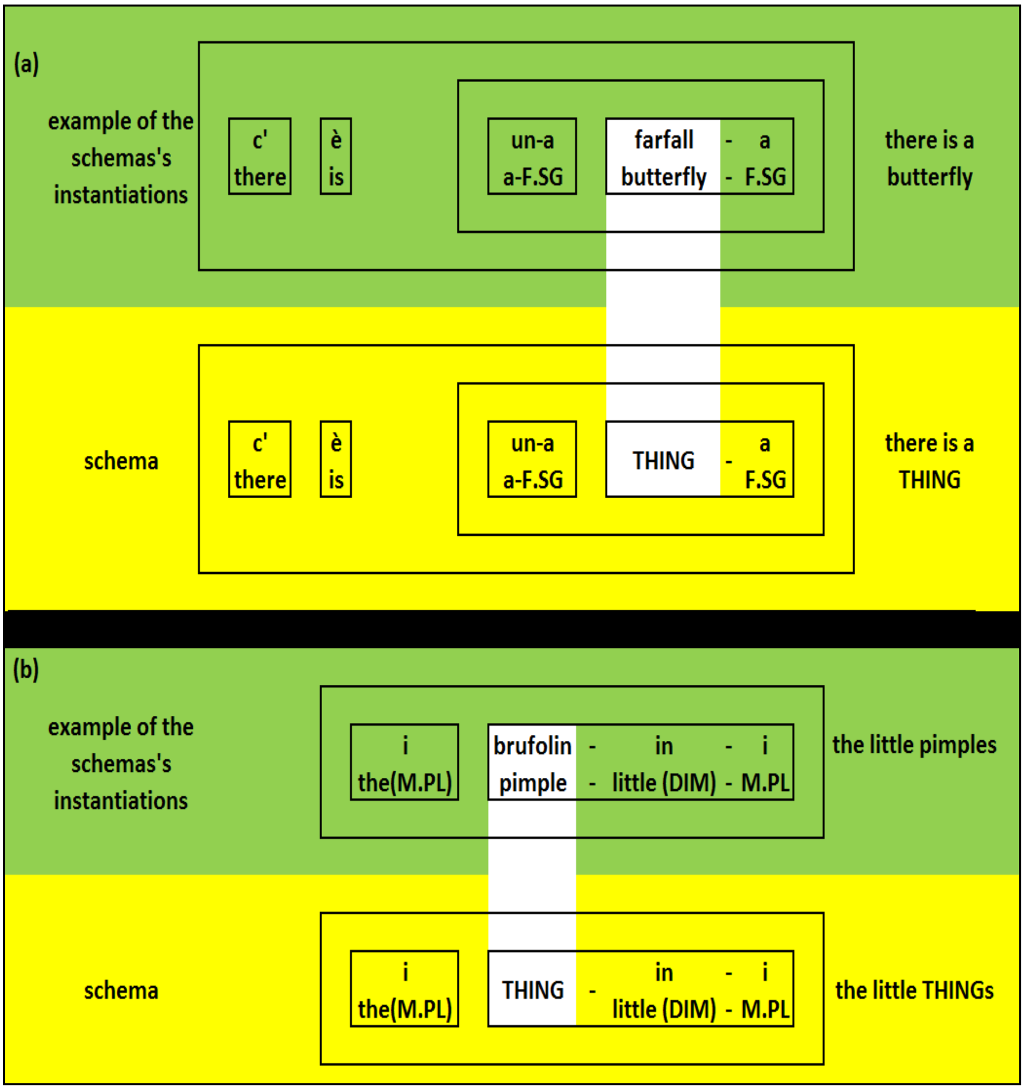


Figure 7.7: the THING slot (highlighted in white); 1 of 2. Schemas are highlighted in yellow and their instantiations are highlighted in green. Slot formation is highlighted in white.

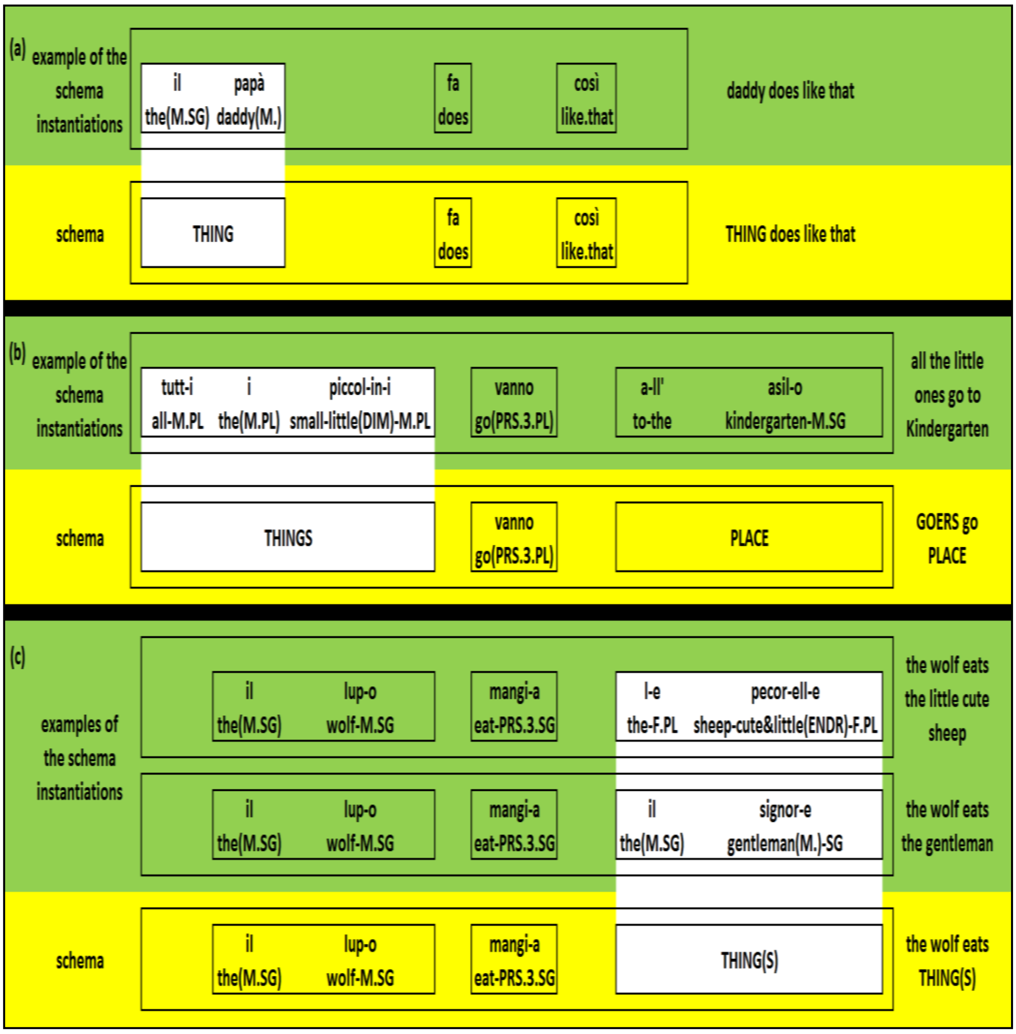


Figure 7.8: the THING slot (highlighted in white); 2 of 2. Refer to fig. 7.7 on how to read this figure.

Table 7.9: types of semantic slots.

Semantic slot type	no.	%
THING	426	54%
PROCESS	230	29%
PLACE	52	7%
SENTENCE	51	6%
QUALITY	14	2%
others	18	2%
total	**791**	**100%**

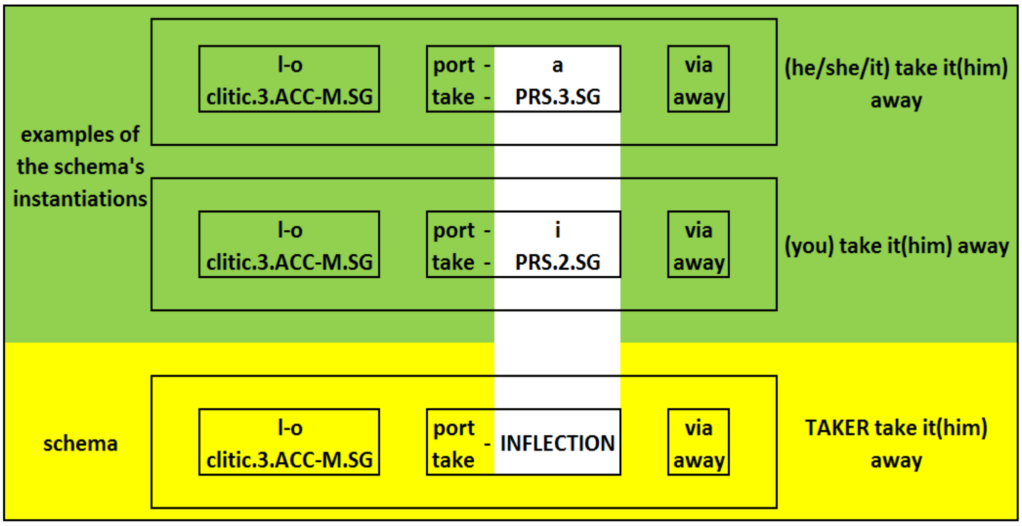


Figure 7.9: process(root)-INFLECTION slot (in white). Refer back to fig. 7.7 on how to read this figure.

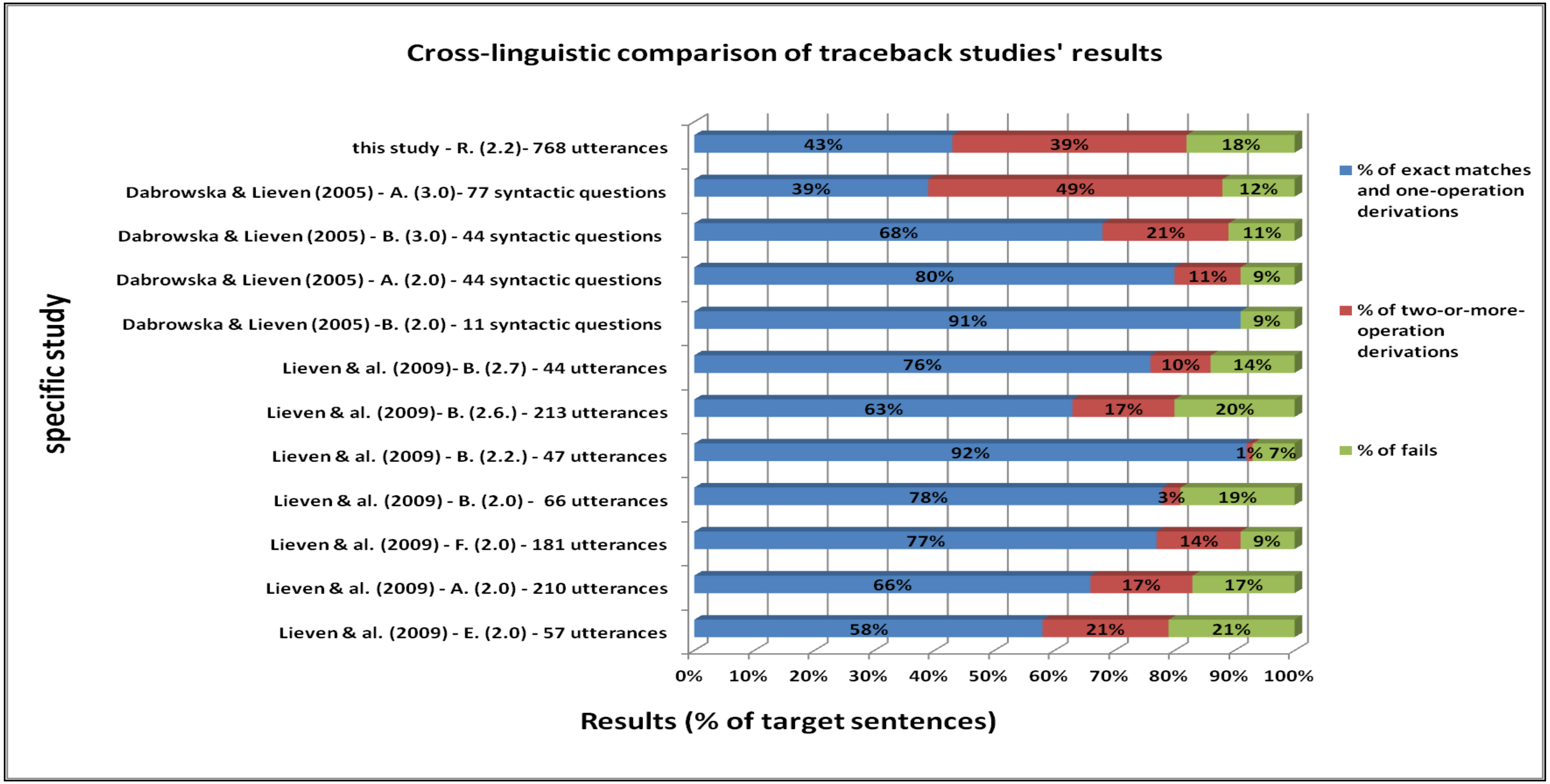


Figure 7.10: Comparing the results of various traceback studies. On the left of each bar it is indicated: the study to which the data belong (year of publication) – child's initial – (child's age) – type and number of constructions enquired.

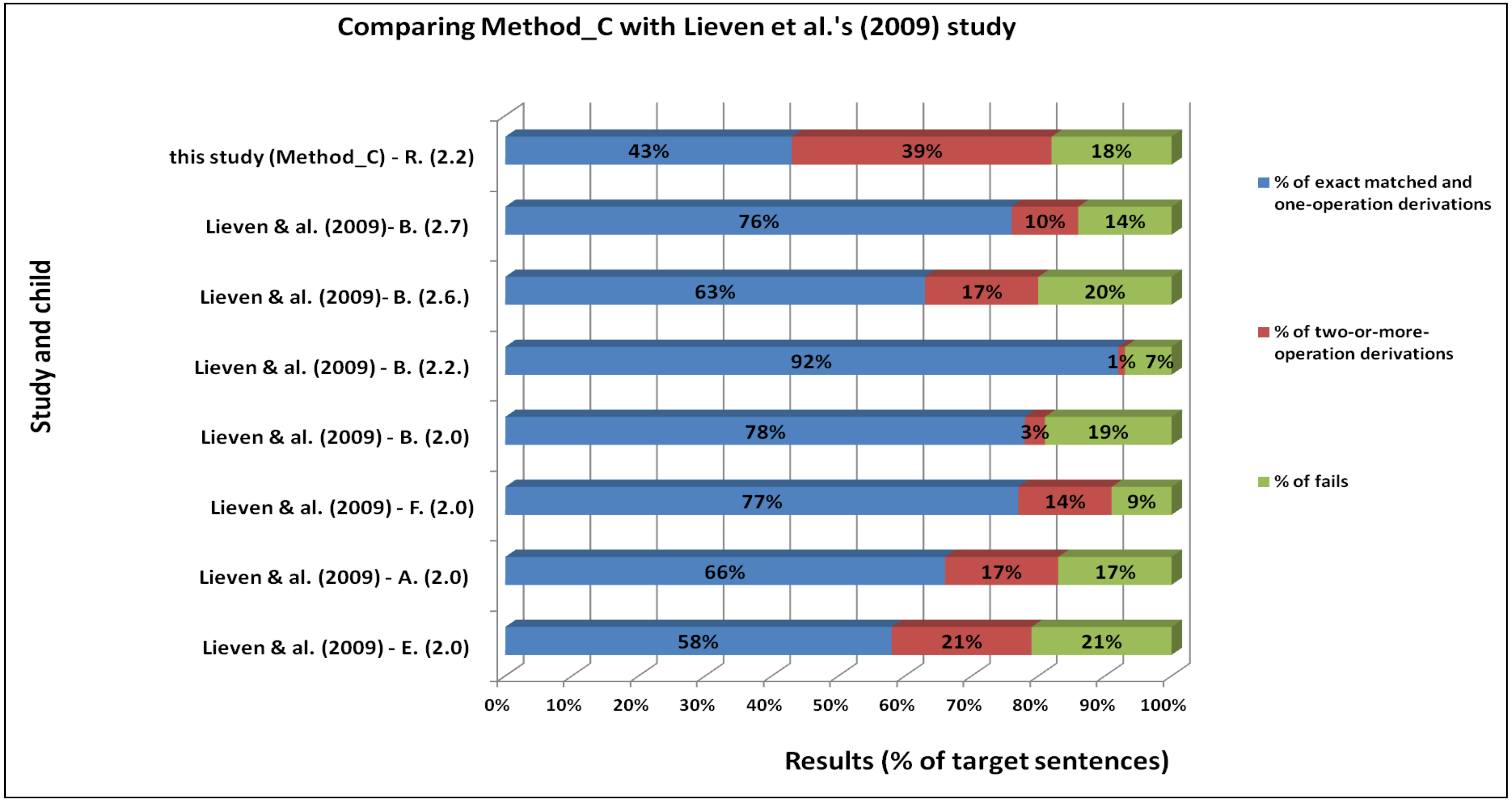


Figure 7.11: Comparing the results of Lieven et al. (2009) with Method_C. On the left of each bar it is indicated: the study to which the data belong (year of publication) – child's initial – (child's age).

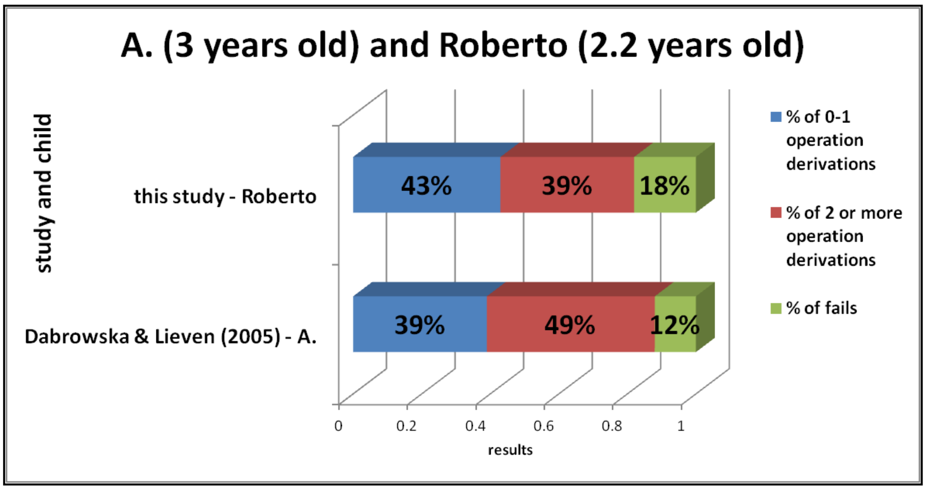


Figure 7.12: Comparing the results of A. (English-speaker, 3;00) and Roberto (Italian-speaker, 2;2).

19.3.4. ANALYSIS

51. ** dai, Luca,* ***scendilo giù!***

da-i,	*Luca,*
give-PRS.2.SG,	Luca,
scend-i=l-o	***giù!***
descend-IMP.2.SG=clitic.3.ACC-M.SG	**down**!

"*Com'on, Luca, descend it/him down!"
"*Com'on, Luca, go it/him down!"
(week6.2014.02.18.B: line 3959)

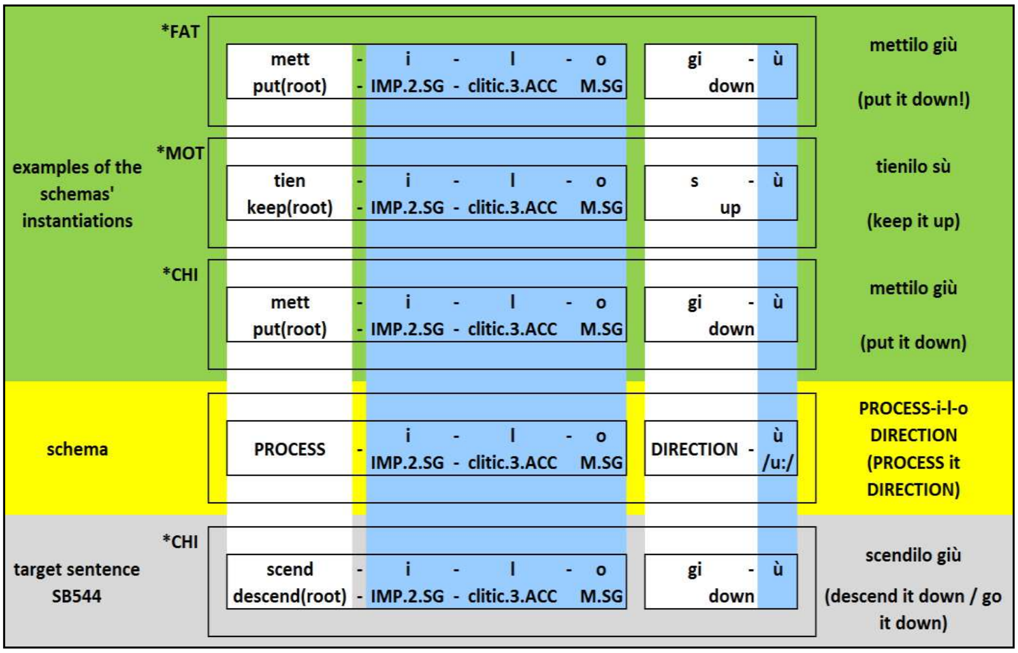


Figure 8.1: The schema (yellow) from which target sentence SB544 (grey) was derived and the schema's instantiations in the main corpus (green). Relationships of elaboration are in white; shared concrete material in blue.

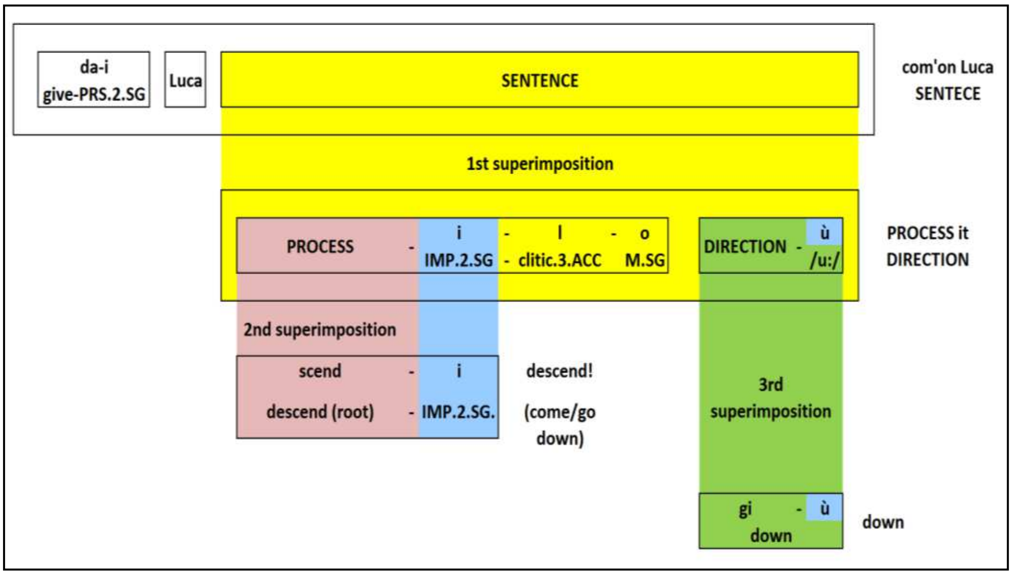


Figure 8.2: deriving target sentence SB544. Slot elaboration is in yellow, green and pink. Shared concrete material is in blue. Please note that the order of the superimpositions/operations in this and other figures is not meant to represent the exact order of assembly. The method makes no assumptions as to the order in which component units are assembled.

52. *CHI: **hai (...) caduto il cagnolino dal trattore .*

hai	*cad-u-to*	*il*
have(PRS.2.SG)	fall-TV(conj.II)-PTCP	the(M.SG)
cagnol-in-o	*da=l*	*trattor-e*
dog-little(DIM)-M.SG	from=the(M.SG)	tractor(M.)-SG

"You have fallen the little dog from the tractor."
(week4.2014.01.28: line 2114)

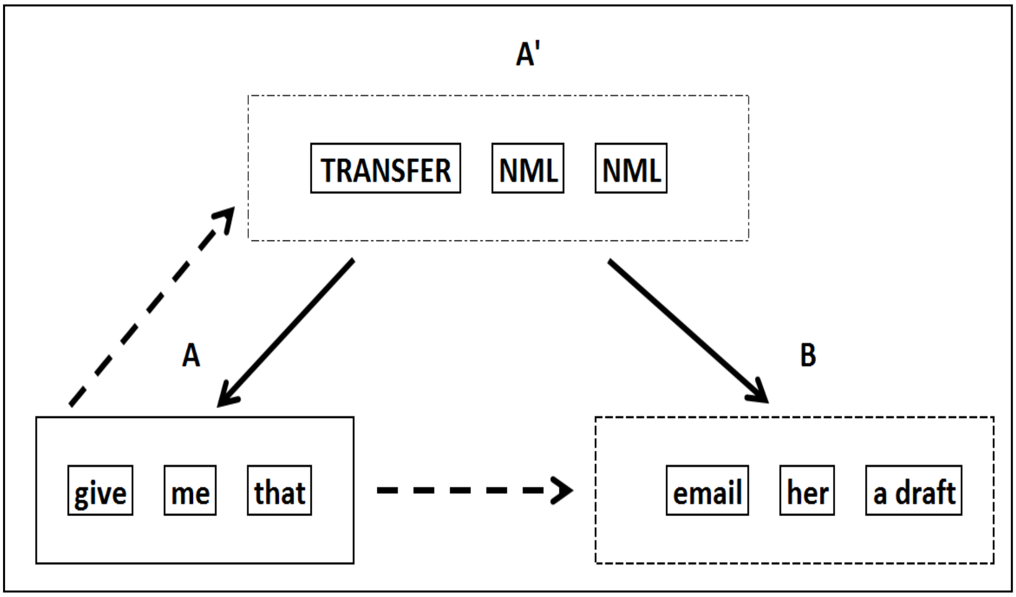


Figure 8.3: Apprehending an instance of the ditransitive construction (B) as an extension vis-à-vis the construction prototype (A). Dashed arrows indicate relationships of extension and solid arrows indicate relationships of elaboration. Thickness of boxes indicates degree of entrenchment.

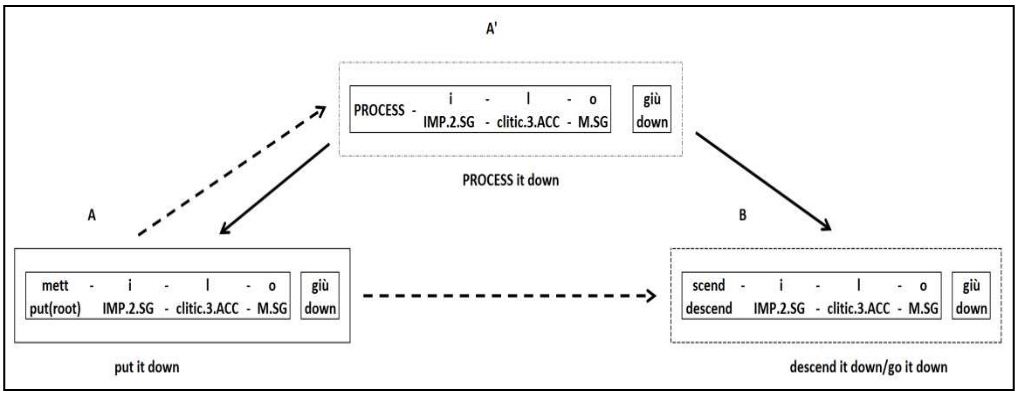


Figure 8.4: Producing target sentence SB544 (B) as an extension vis-à-vis a prototypical instantiation of the cmc (A). See figure 8.3 on how to interpret lines, arrows and boxes.

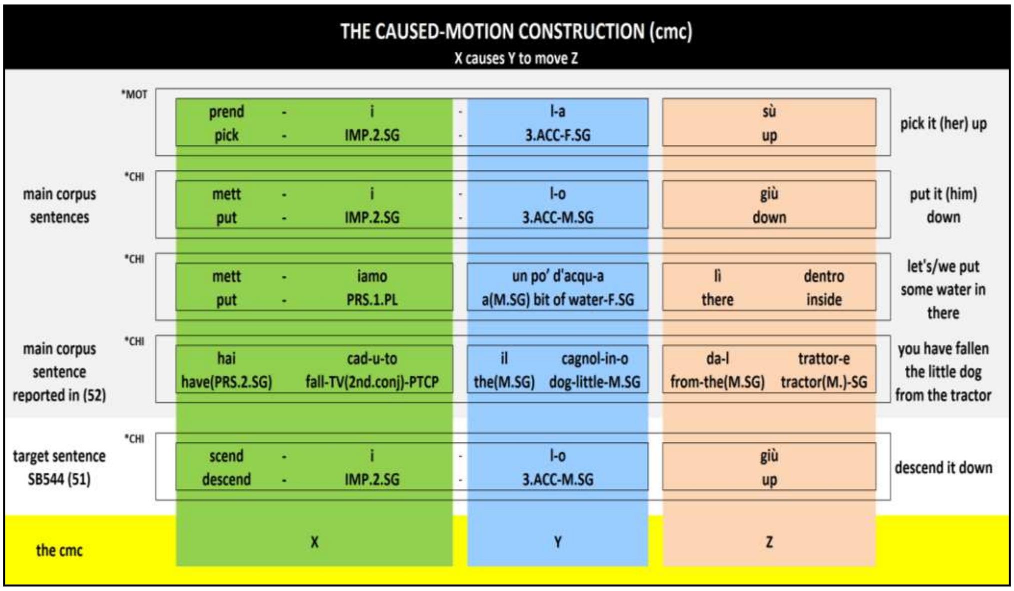


Figure 8.5: the caused motion construction and (some of) its instantiations in the corpus collected.

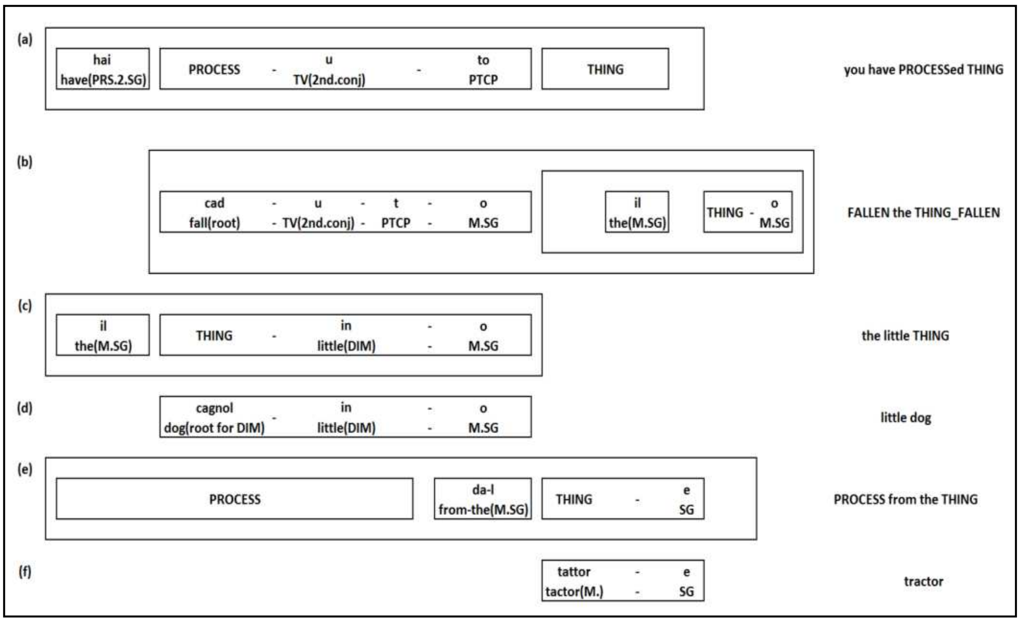


Figure 8.6: the units to which the sentence in (52) is traced back.

Figure 8.7: deriving the sentence in (52). Slot elaboration is highlighted in yellow and shared lexical material is highlighted in blue.

Table 8.1: Distribution of morphological fails.

fail type	no.	% of 15 morphological fails	% of 135 fails	% of 768 target sentences
morphological fails	11	74%	8%	1.5%
both syntactic and morphological fails	2	13%	2%	0.3%
both morphological and lexical fails	2	13%	1%	0.3%
total	**15**	**1**	**0.11**	**2.10%**

53. *I cervi vanno a nanna*

i	*cerv-i*	*vanno*	*a*	*nann-a*
the(M.PL)	deer-M.PL	go(PRS.3.PL)	to	beddy.bye-F.SG

"The deer go to sleep."
(week6.2014.02.18.A: line 719-720)

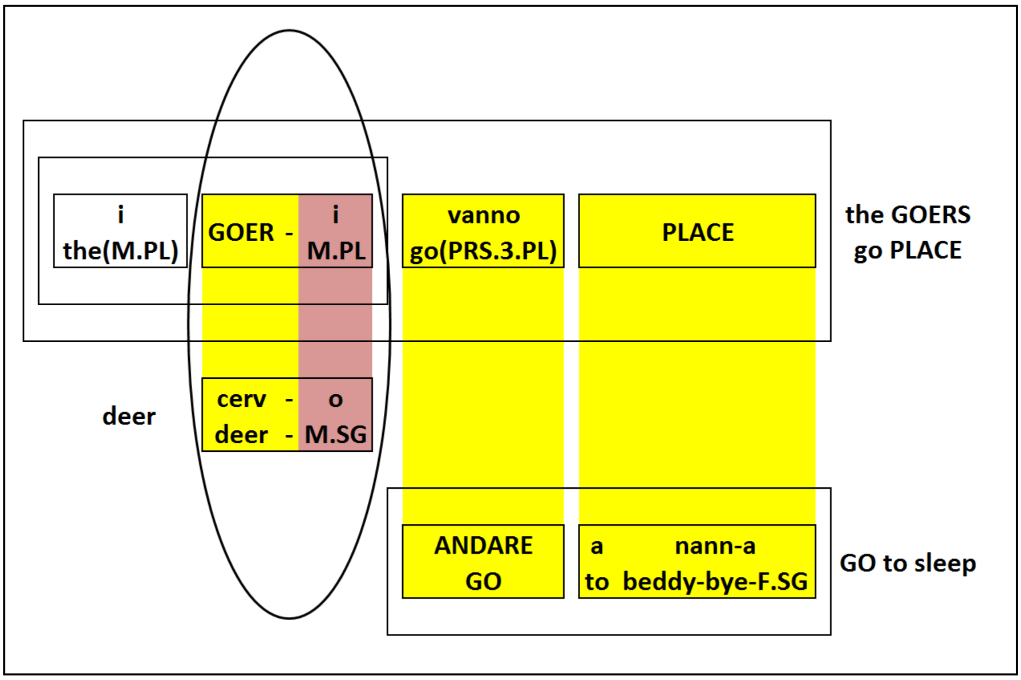


Figure 8.8: failing to derive part of target sentence SA079 (53). Slot elaboration is in yellow. Contrasting morphological specifications are highlighted in red.

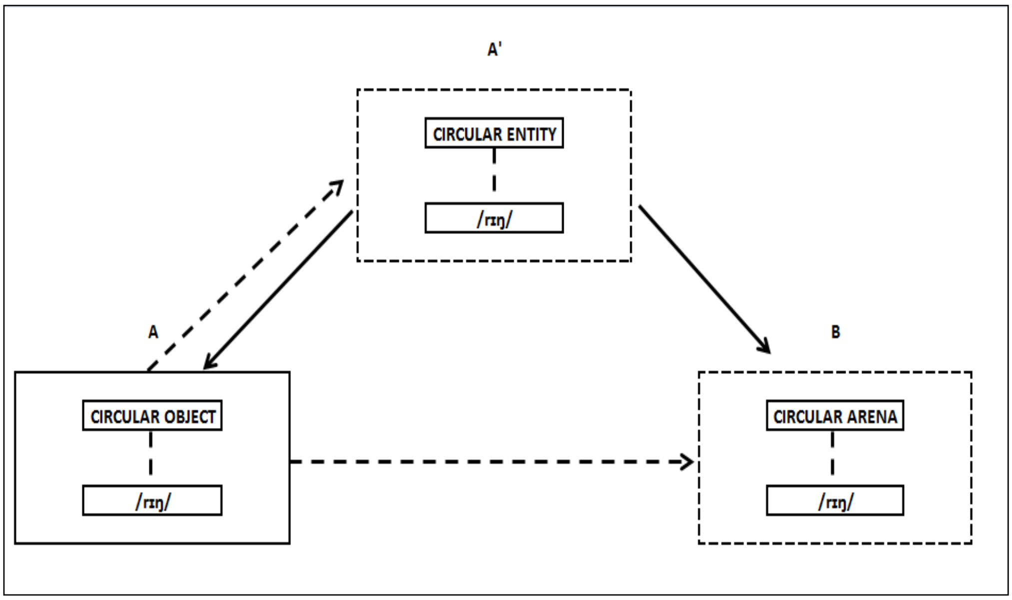


Figure 8.9 categorising B as an extension from A (from Langacker, 2000, 2008). See fig. 8.3 on how to read this figure.

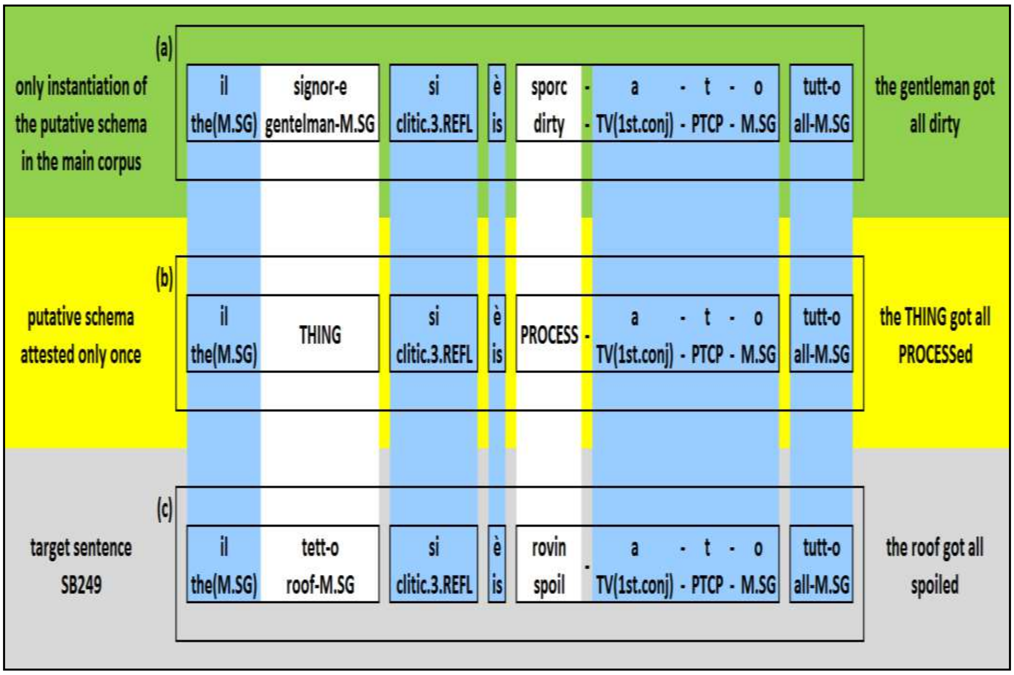


Figure 8.10: deriving target sentence SB249 (c) from its component units under Method_C and Method_D; 1 of 2. Slot formation is highlighted in white and recurring lexical material is highlighted in blue.

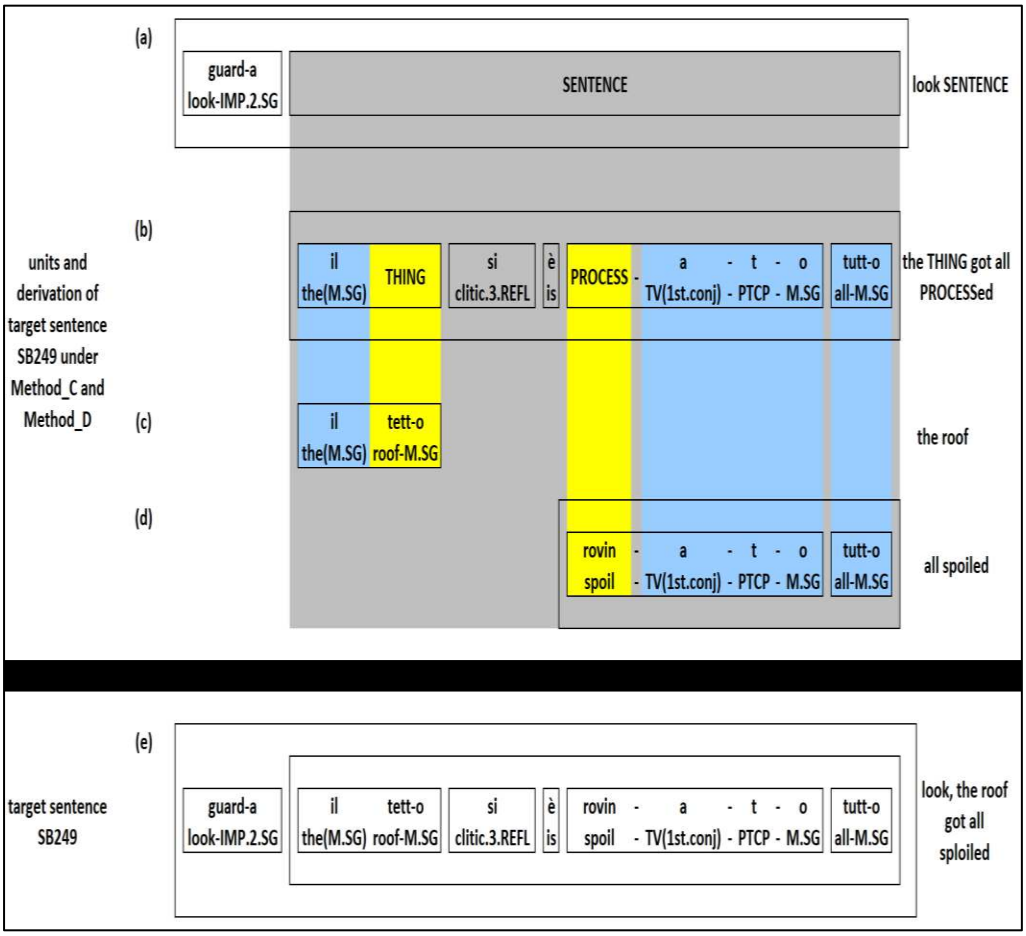


Figure 8.11: deriving target sentence SB249 (e) from its component units (a-d) under Method_C and Method_D; 2 of 2. Superimpositions are highlighted in grey and yellow. Shared lexical material is highlighted in blue.

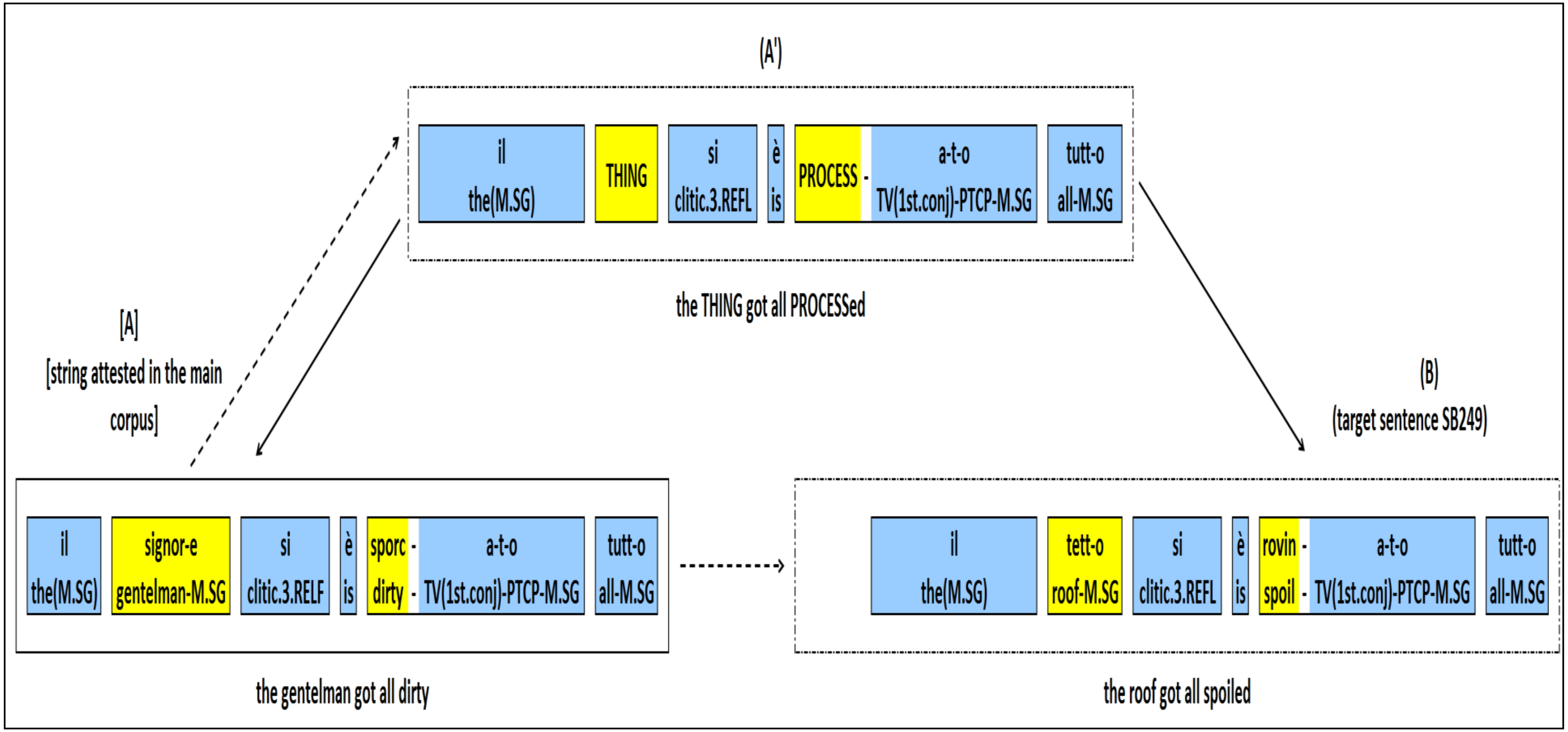


Figure 8.12: target sentence SB249 (B) as an extension vis-à-vis a string attested in the main corpus (A). Strings in thick boxes have status of units; strings in dashed boxes do not have status of unit. Solid arrows are relationships of elaboration; dashed arrows are relationships of extension. Words highlighted in blue indicate shared lexical material, the material in yellow indicate relationships of elaboration or extension.

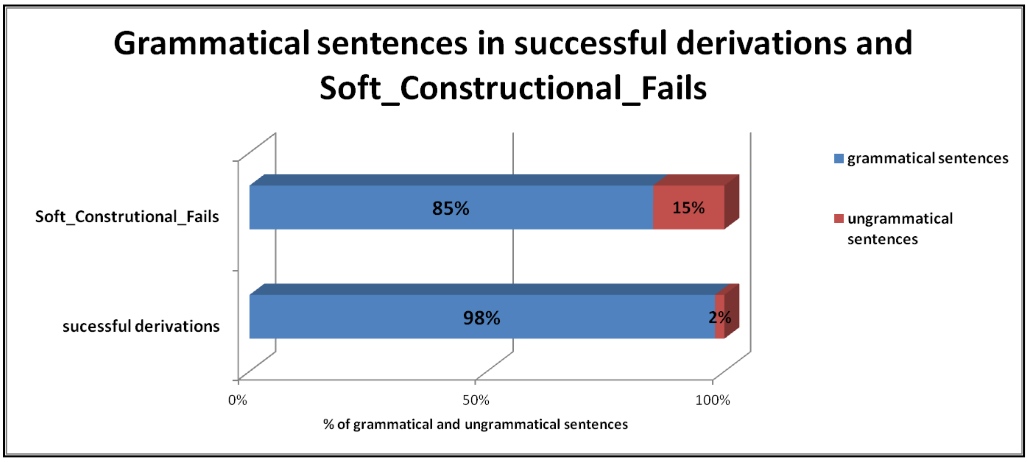


Figure 8.13 grammatical and ungrammatical sentences in successful derivations and Soft_Constructional_Fails.

54. *La mia nonna mi ha dato questo piatto*
l-a mi-a nonn-a mi ha
the-F.SG my-F.SG grandma-F.SG clitic.1SG.DAT has
d-a-to quest-o piatt-o
give-TV(conj.I)-PTCP this-M.SG dish-M.SG
"My grandma gave me this dish."
(week.6.2014.02.18.14; line 1155)

55. *La nonna mi ha dato un boccon-e grand-e*
l-a nonn-a mi ha
the-F.SG grandma-F.SG clitic.1.SG.DAT has
d-a-to un boccon-e grand-e
give-TV(conj.I)-PTCP a(M.SG) mouthful(M.)-SG big-SG
"Grandma gave me a big mouthful."
(week.6.2014.02.18.14; line 1155)

56. *Era molto scarica la moto*
era molto scaric-a l-a mot-o
was very out.of.battery-F.SG the-F.SG motorbike(F.)-SG
"The motorbike was really out of battery."
(week6.2014.02.18.B: line 1491)

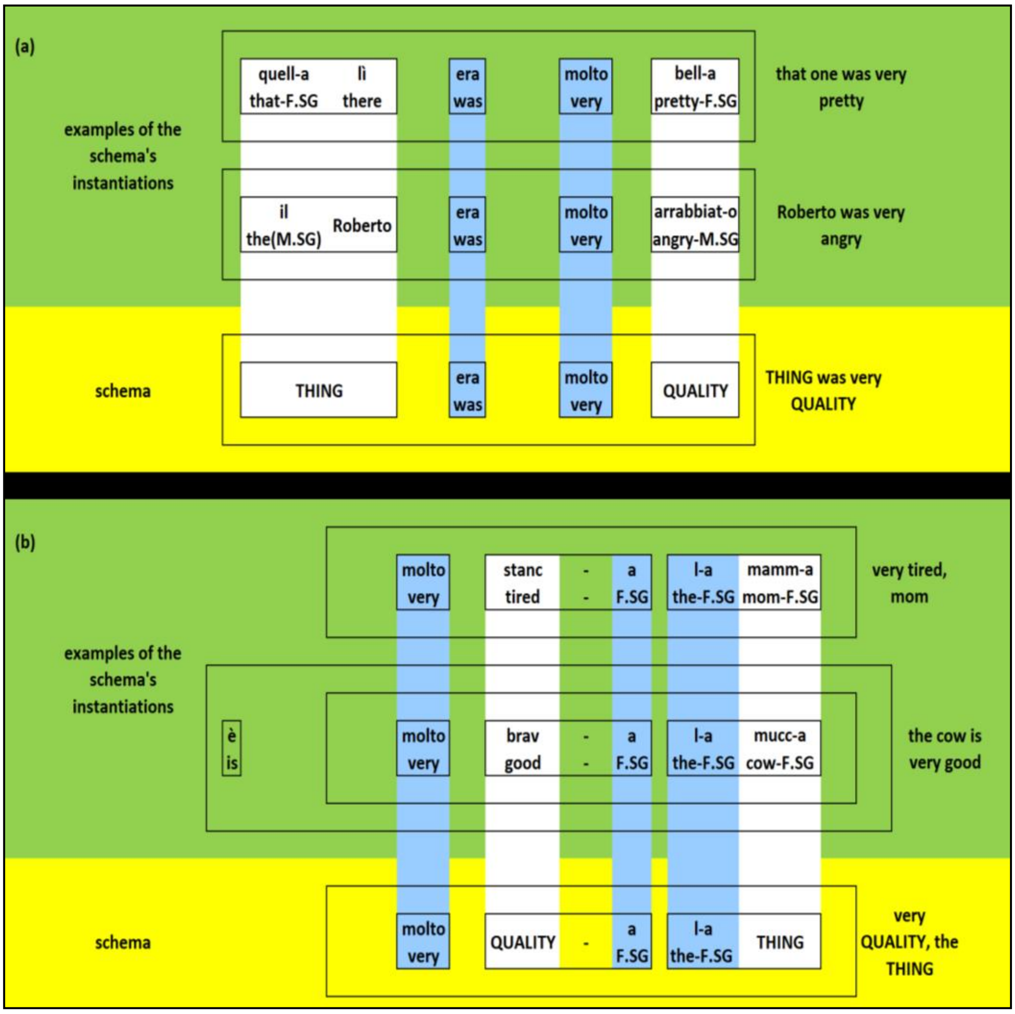


Figure 8.14 the precedents of target sentence SB192 (56). Schemas are in the yellow strips and their instantiations in the green strips. Relationships of elaboration are in white and shared concrete material is in blue.

Table 8.2: distribution of Hard_Constructional_Fails.

Type of Hard_Constructional_Fail	no. of sentences	% of 50 hcf	% of 768 target sentences
successful derivations that do not account for some kind of agreement within the sentence	13	26%	2%
accountable for by concatenations of partial overlaps	12	24%	1.6%
multiple issues	10	20%	1.3%
phonological mistake, element dropped from an attested schema and ill-formed imitation of the preceeding sentence	9	18%	1%
accountable for by insertion of a component unit within another component unit	5	10%	1%
explainable through extension and the co-activation of two different schemas	1	2%	0.1%
total	**50**	**100%**	**7%**

57. a) *RES: *io c' **ho la testa sottosopra**.*
io c(i) ***ho*** ***l-a*** ***test-a***
I clitic.DAT **have(PRS.1.SG) the-F.SG head-F.SG**
sotto+sopra
under+over
"I've got my head upside down."

b) *CHI: *hai la testa sopra.*
hai *l-a* *test-a* *sopra*
have(PRS.2.SG) the-F.SG head-F.SG over/up
"You've got the head over/up."

(week6.2014.02.18.B: lines 3098 and 3099)

Table 8.3: distribution of ungrammatical sentences.

type of target sentence	no. of ungrammatical sentences	% of 41 ungrammatical sentences
successful derivations	11	27%
lexical fails	1	2%
Soft_Constructional_Fails (scf)	7	17%
Hard_Constructional_Fails (hcf)	14	34%
Phonological mistakes, elements dropped from an attested schema and ill-formed imitations.	8	20%
total	41	100%

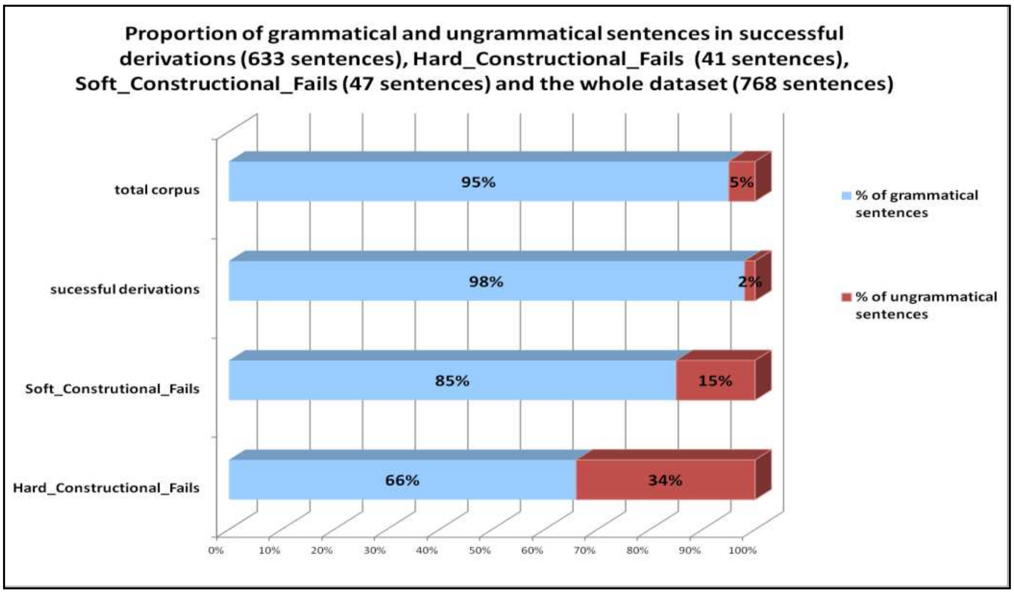


Figure 8.15: proportion of grammatical and ungrammatical sentences in successful derivations, Hard_Constructional_Fails, Soft_Constructional_Fails and the whole dataset.

Table 8.4: comparing grammatical vs. ungrammatical sentences in Successful Derivations, Hard_Constructional_Fails and Soft_Constructional_Fails. The table presents odds ratios (and CIs) for each comparison.

	Successful Derivations	Soft_Constructional_Fails	Hard_Constructional_Fails
Successful Derivations			
Soft_Constructional_Fails	9.82** (3.05 - 29.55)		
Hard_Constructional_Fails	28.82** 11.04 - 77.77	2.93 0.95 - 9.77	

** p < .05 (Bonferroni correction)

58. *CHI: *tu <sei> [/] <sei> [/] sei come (..) da tagliare la carne, Luca.*
tu sei come da tagliare la carne, Luca
tu *sei* *come* *(..)* *da*
you(2.SG.NOM) be(PRS.2.SG) like (..) to
tagli-a-re *l-a* *carn-e* *Luca*
cut-TV(conj.I)-INF the-F.SG meat(F.)-SG Luca
"Luca, you are like (..) to cut the meat."
(week6.2014.18.B: lines 3372-3373)

59. *CHI: **devi dare il gistratore [: registratore] [* p] a qualcuna [: qualche] [*] persona.*
devi dare il registratore a qualcuna persona.
dev-i *d-a-re* *il*
have.to-PRS.2.SG give-TV(conj.I)-INF the(M.SG)
registrator-e *a* *qualcun-a* *person-a.*
voice.recorder(M.)-SG to someone-F.SG person-F.SG
("*you have to give the voice recorder to someone person").
"You have to give the voice recorder to someone."
(week.6.2014.02.18.B: line 3324-3325)

60. a) qualche THING-a.
some THING-F.SG.

b) *qualcun-GENDER.NUMBER.*
someone-GENDER.NUMBER.

61. *CHI: **<respondiamo [: rispondiamo] [* p] > [<] a qualcuno [: qualche] [* s:r]> [?] bambido [: bambino] [* p]*
rispondiamo a qualcuno bambino
rispond-iamo *a* *qualcun-o* *bambin-o*
answer-PRS.1.PL to someone-M.SG child-M.SG
("*we answer to someone child").
"We answer to some child."
(week5.2014.02.04: lines 1389-1390)

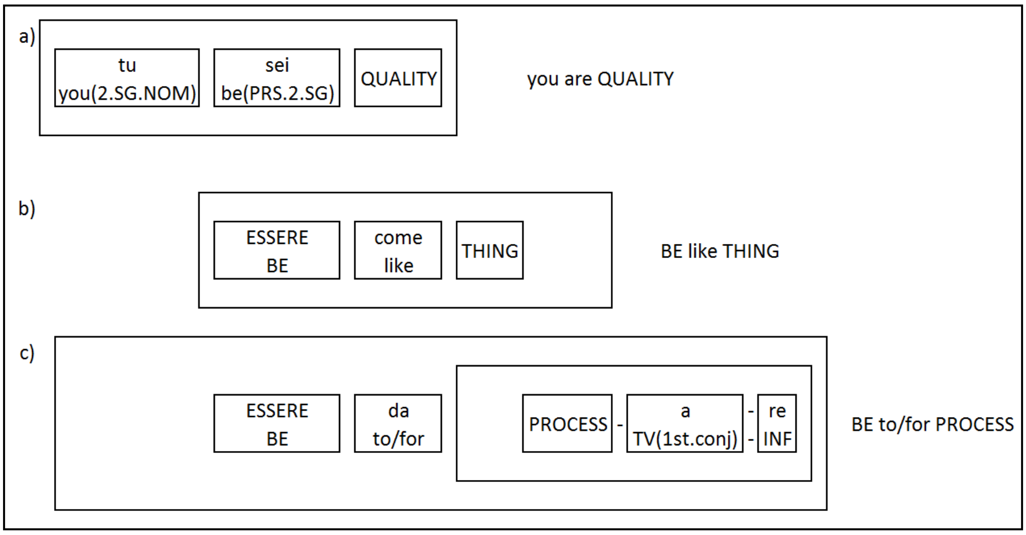


Figure 8.16: (some of) the putative precedents of target sentence SB473 (58).

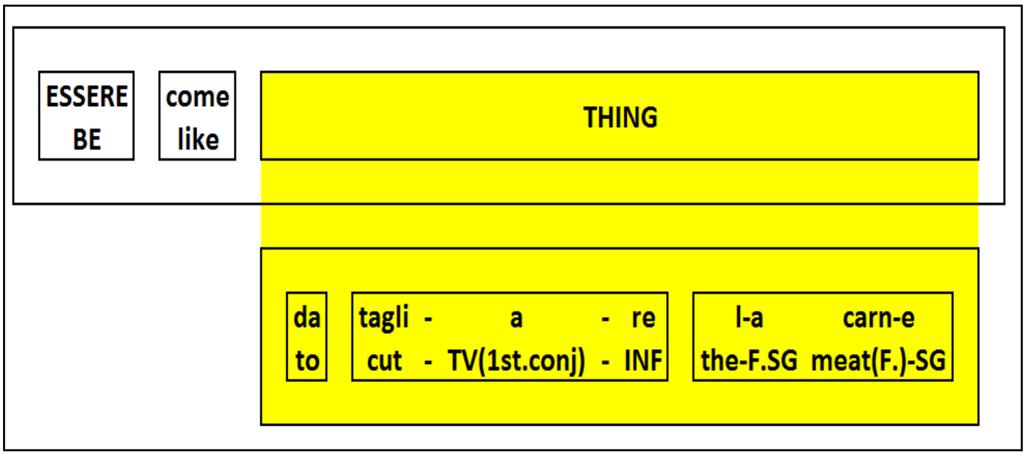


Figure 8.17: elaborating a *THING* slot with *da tagliare la carne*. The superimposition is in yellow.

62. [*da mangiare*] - - - ➔ (*da tagliare la carne*).

63. da PROCESS-TV-re.
to PROCESS-TV-INF.
"To PROCESS/ for PROCESSing."

64. a) *RES: *il tacchino è buono da mangiare*
il *tacchin-o* *è* *buon-o*
the(M.SG) turkey-M.SG is good-M.SG
da *mangi-a-re*
to eat-TV(conj.I)-INF
"The turkey is good to eat."
(week2.2014.01.14.A: line 384)

b) *MOT: *serve per dare **da bere** ai fiorellini .*
serv-e *per* *d-a-re*
be.for-PRS.3.SG to give-TV(conj.I)-INF
da ***b-e-re*** *a=i*
to **drink-TV(conj.II)-INF** to=the(M.PL)
fiorell-in-i
flower-little(DIM)-M.PL
("(It) is for giving the little flowers **(something) to drink**").
"It is for watering the flowers."
(week2.2014.01.20: line 423)

c) *CHI: *ti do **da mangiare***
ti *d-o* ***da*** ***mangi-a-re***
clitic.2.SG.DAT give-PRS.1.SG **to** **eat-TV(conj.I)-INF**
"(I) give you (**something) to eat**."
(week4.2014.01.31: line 932)

65. *Sputare non è bello*
sput-a-re *non* *è* *bell-o*
spit-TV(conj.I)-INF not is nice-M.SG
"Spitting is not nice."

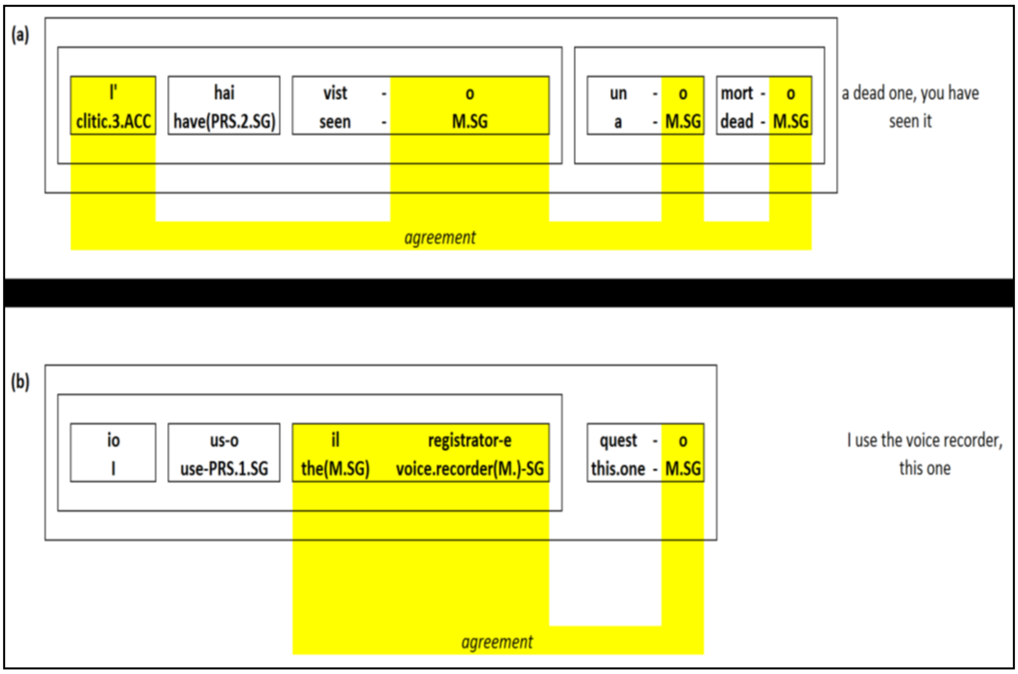


Figure 8.18: agreement between object-NML and resumptive element. Correct agreement is in yellow.

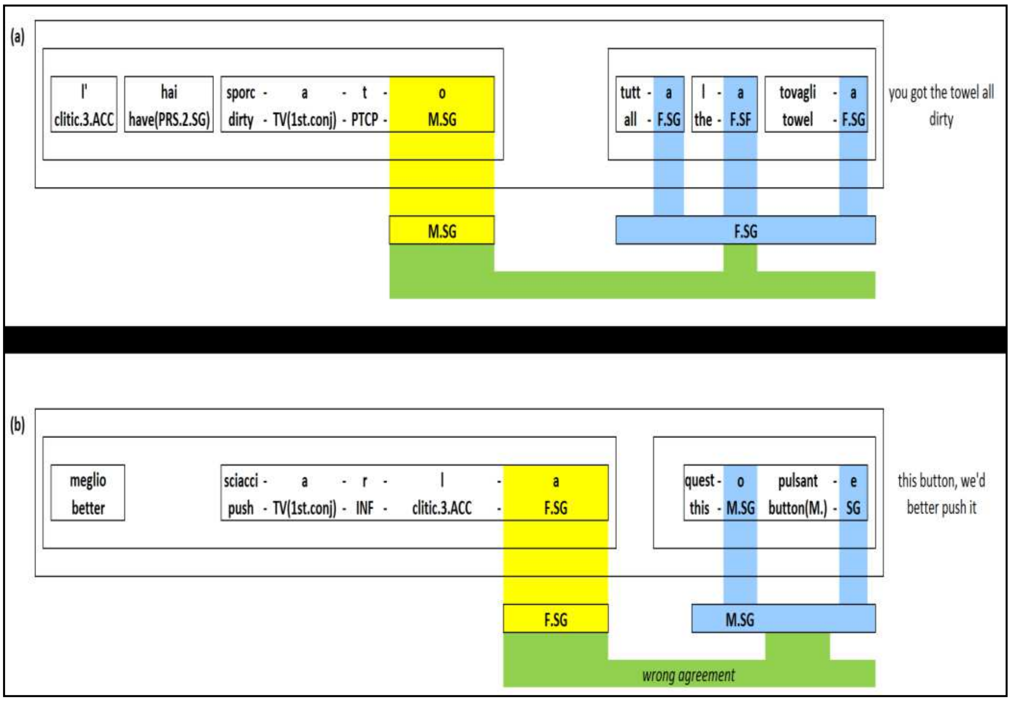


Figure 8.19: agreement between object-NML and resumptive element. Wrong agreement is in green.

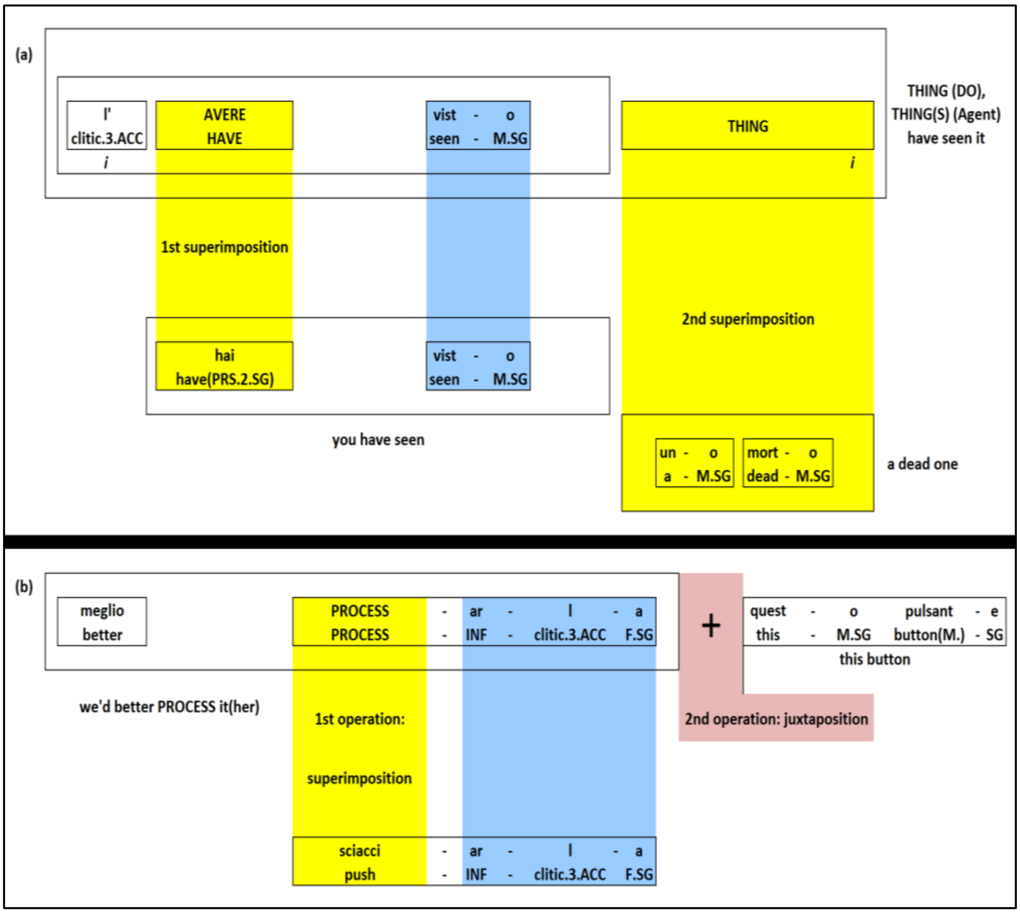


Figure 8.20: Deriving fig. 8.18a (a) and fig. 8.19b (b). Slot elaboration is highlighted in yellow and shared lexical material is in blue. The red highlighting and the sign plus (+) indicate the juxtaposition of units.

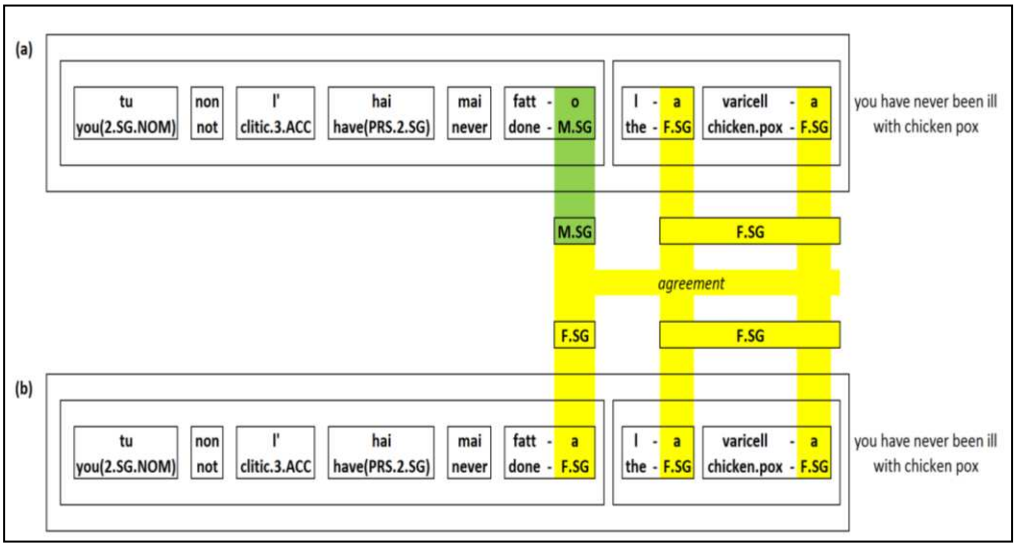


Figure 8.21: when adults deliver wrong agreement (highlighted).

66. *CHI: *questo omino è rimasto lì dentro*

*quest-**o***	*om-in-**o***	*è*	*rima-st-**o***
this-**M.SG**	man-little(DIM)-**M.SG**	is	stay-PTCP-**M.SG**
lì	*dentro*		
there	inside		

"This little man stayed in there."
(week6.2014.12.18.A: line 988)

67. *CHI: *gli stivaletti sono tutti sporchi*

gli	*stival-ett-**i***	*sono*	*tutt-**i***
the(M.PL)	boot-little(DIM)-**M.PL**	be(PRS.3.PL)	all-**M.PL**
*sporch-**i***			
dirty-**M.PL**			

"The little boots are all dirty."
(week6.2014.02.18.A: line 1100)

68. *CHI: *quello non è uno uovo*[2]

*quell-**o***	*non*	*è*	*un-**o***	*uov-**o***
that-**M.SG**	not	is	a-**M.SG**	egg-**M.SG**

"That is not an egg."
(week6.2014.02.14.B: line 1662)

[2] Technically, in this sentence Roberto uses the wrong article, *uno uovo* instead of *un uovo*. *Un* should be used when the NP starts with a vowel. However, since both *un* and *uno* are indefinite masculine articles, the sentence is considered grammatical as the child appropriately chooses the M.SG article.

69. *CHI: *quando c' era (.) il papà &*MOT:sì*[3] *che faceva (.) la pizza.*

quando	*c(i)*	*er-a*	*il*
when	there(clitic.LOC)	be(IMPERF)-3.SG	the(M.SG)
papà	*che*	*fac-ev-a*	*l-a*
daddy(M.)	that	make-IMPERF-3.SG	the-F.SG
pizz-a.			
pizza-F.SG			

"When there was daddy making pizza."

(week6.2014.02.18: line 683)

[3] Please note that, according to CHAT conventions (MacWhinney, 2000), the insertion of "&*MOT" means that the mother just says the word *sì* "yes", without interrupting the child's speech. That is, the mother encourages the child to continue, which he does by adding the relative clause to the main one.

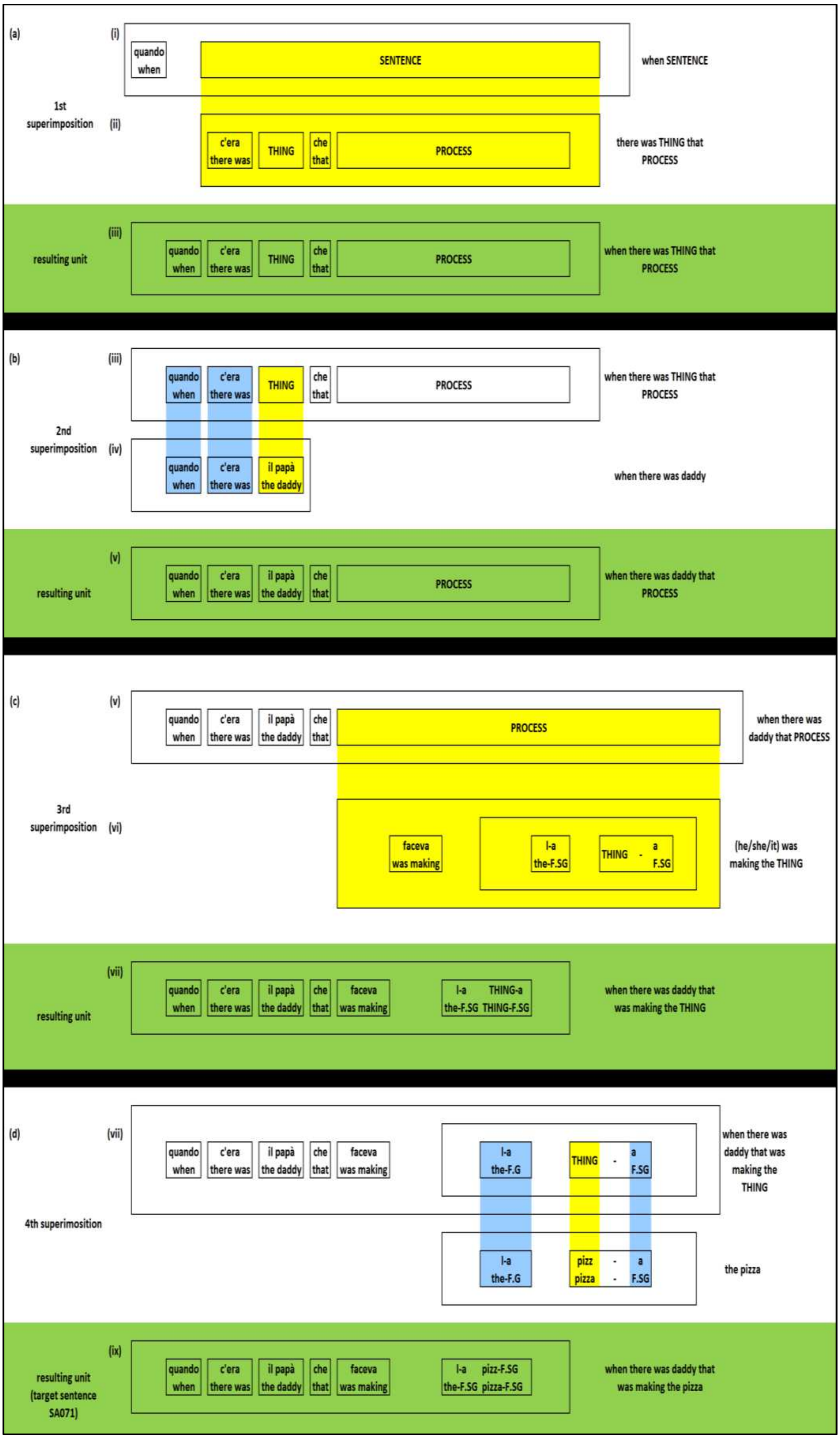


Figure 8.22: deriving target sentence SA071 (69). 8.22c is the problematic superimposition that causes the fail. Slot elaboration is highlighted in white and shared lexical material is highlighted in blue.

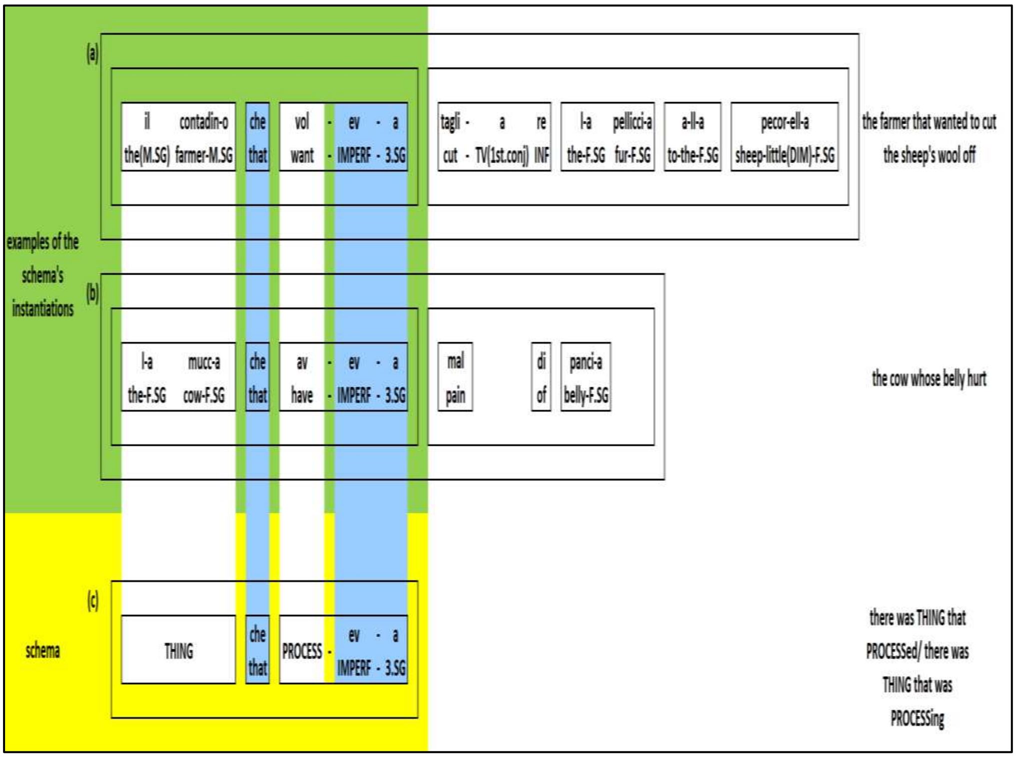


Figure 8.23: the schema (in yellow) that could account for subject-verb agreement in target sentence SA071 (69) and it instantiations (in green). The parts with white backgrounds do not contribute to the creation of the schema.

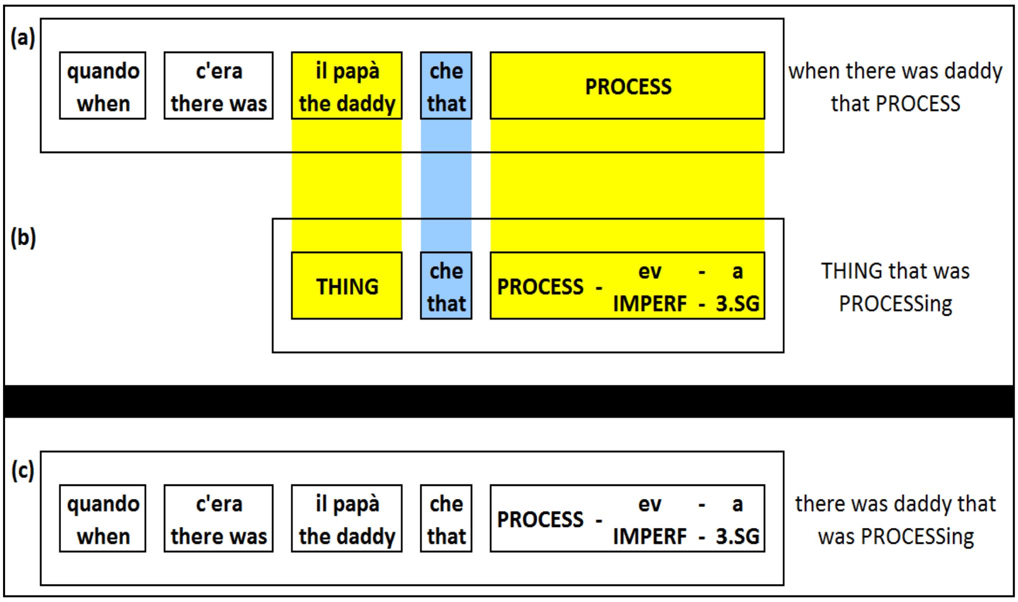


Figure 8.24: explaining subject-verb agreement in target sentence SA071 (69). The unit in c results from the superimposition of a and b. Slot elaboration is in yellow and overlapping (shared) lexical material is highlighted in blue.

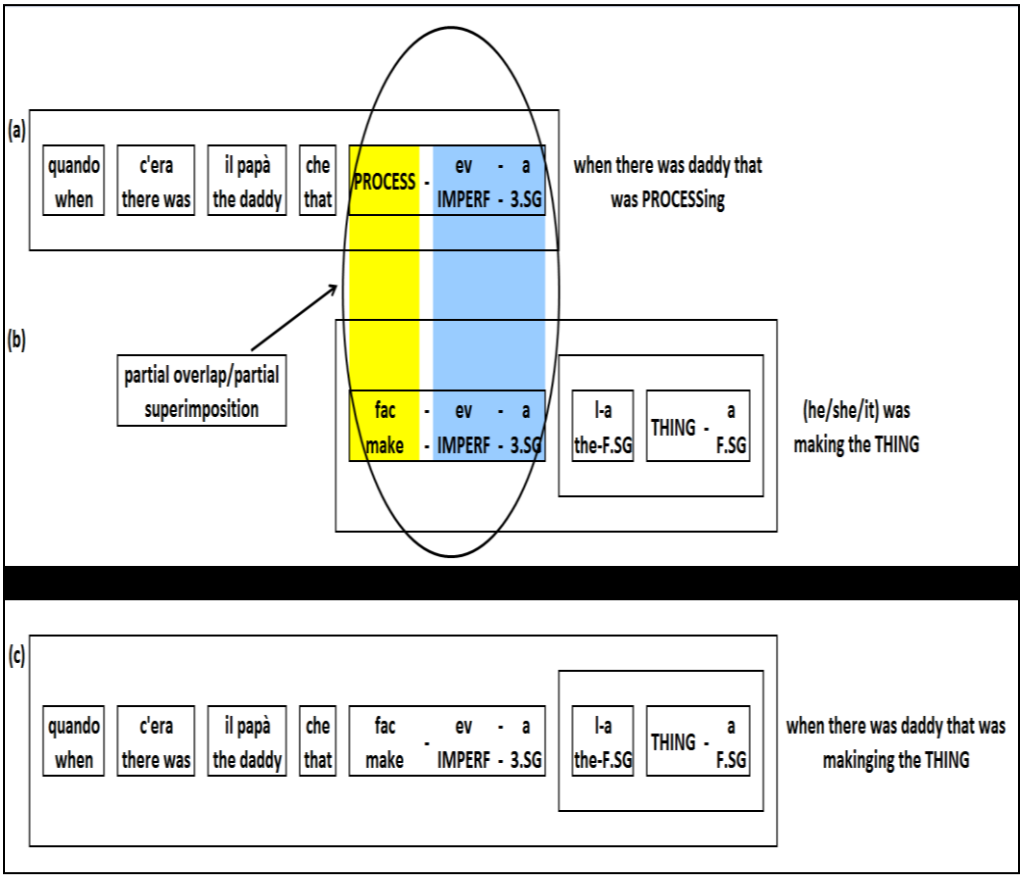


Figure 8.25: explaining target sentence SA071 (69) through partial overlap. The unit in c results from the partial overlap of a and b. See previous figure on how to interpret the colours.

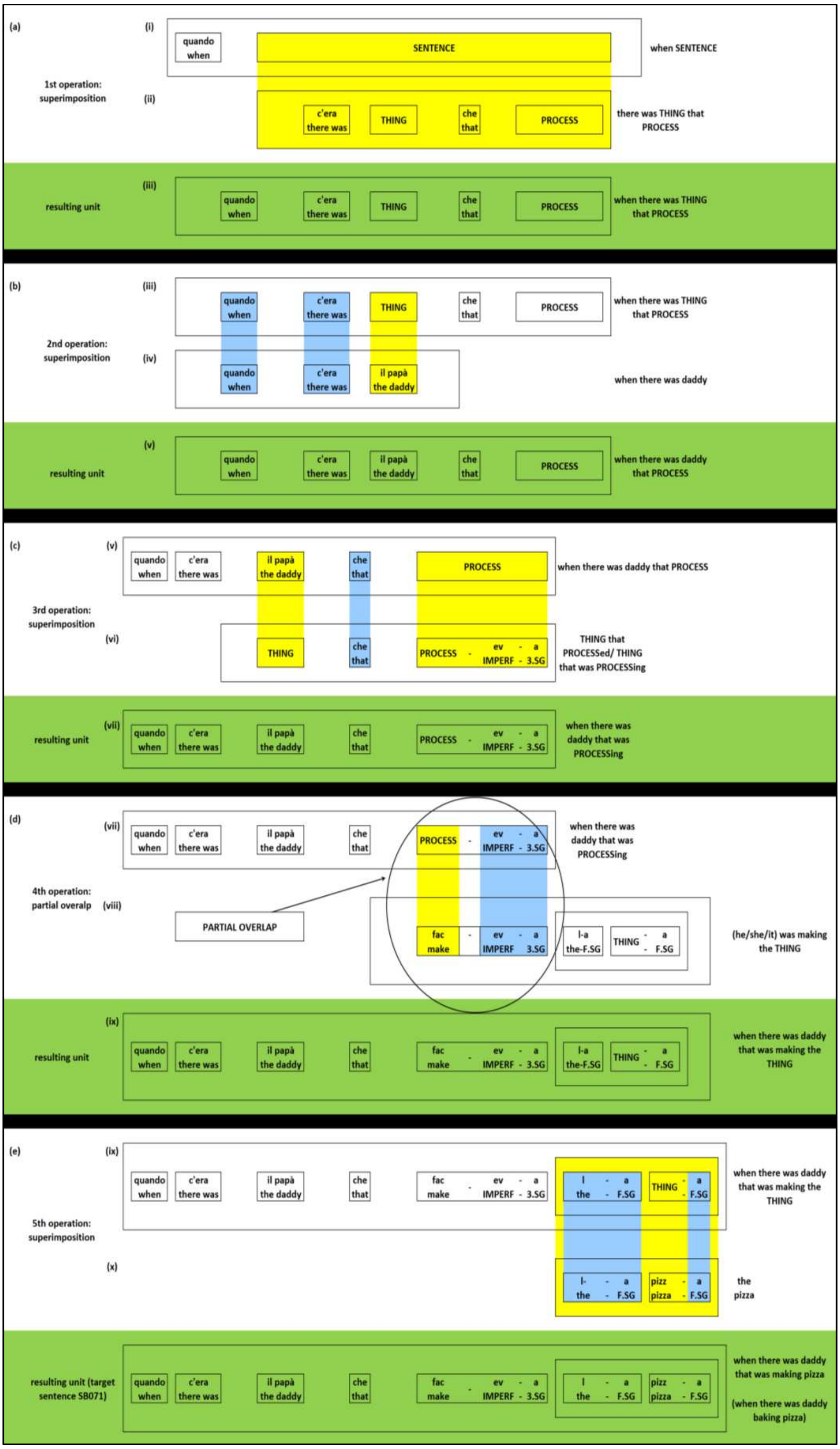


Figure 8.26: deriving target sentence SA071 (69) through partial overlap. Slot elaboration is highlighted in yellow. Shared overlapping material is in blue.

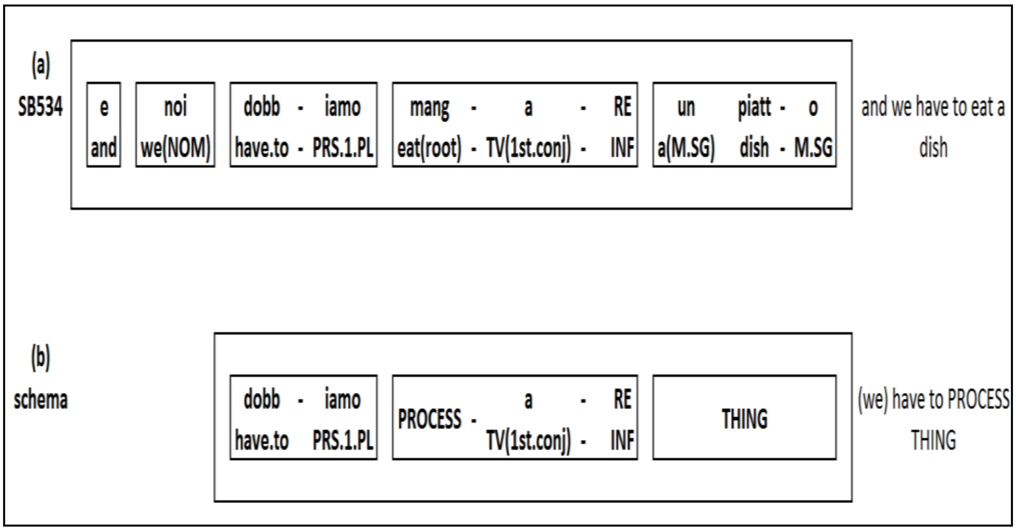


Figure 8.27: target sentence SB534 (a) and the subjectless schema that could account for it (b).

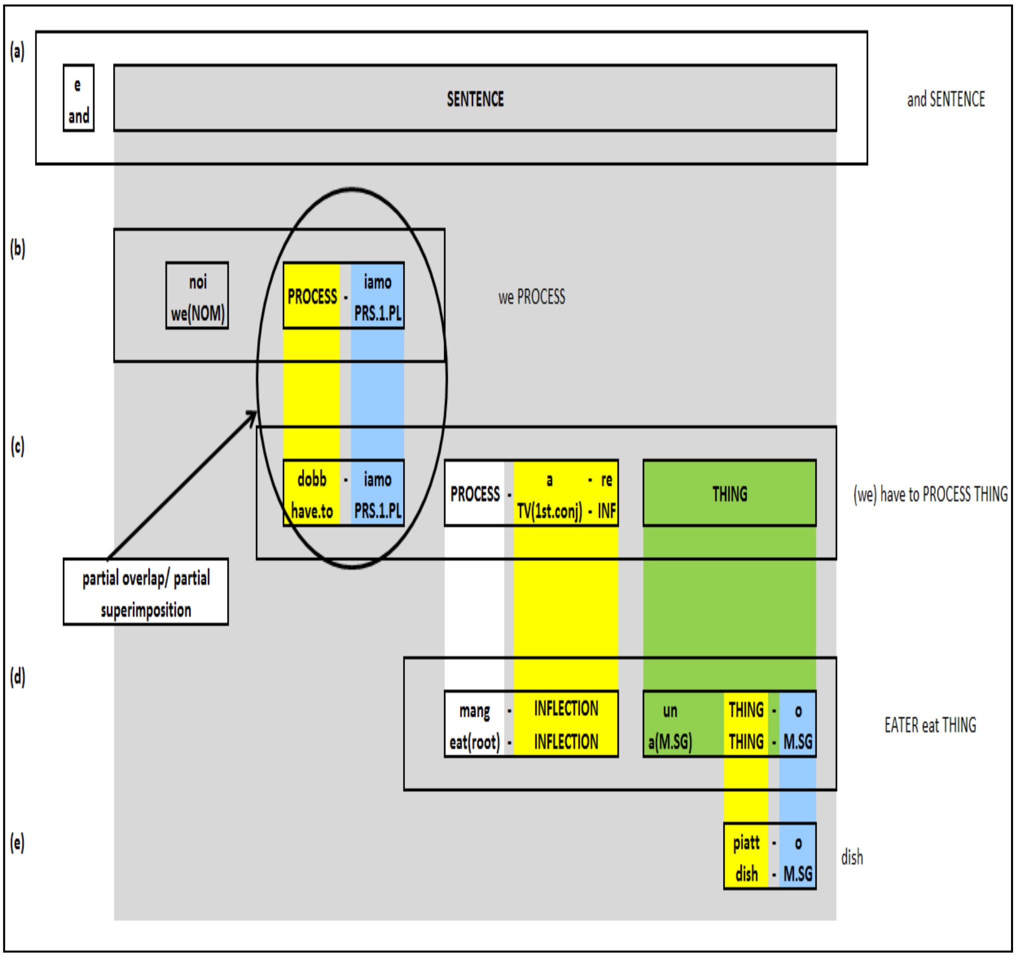


Figure 8.28: deriving target sentence SB534 through partial overlap (in the circle). Elaboration relationships are highlighted in grey, white, yellow and green. Shared concrete material is highlighted in blue.

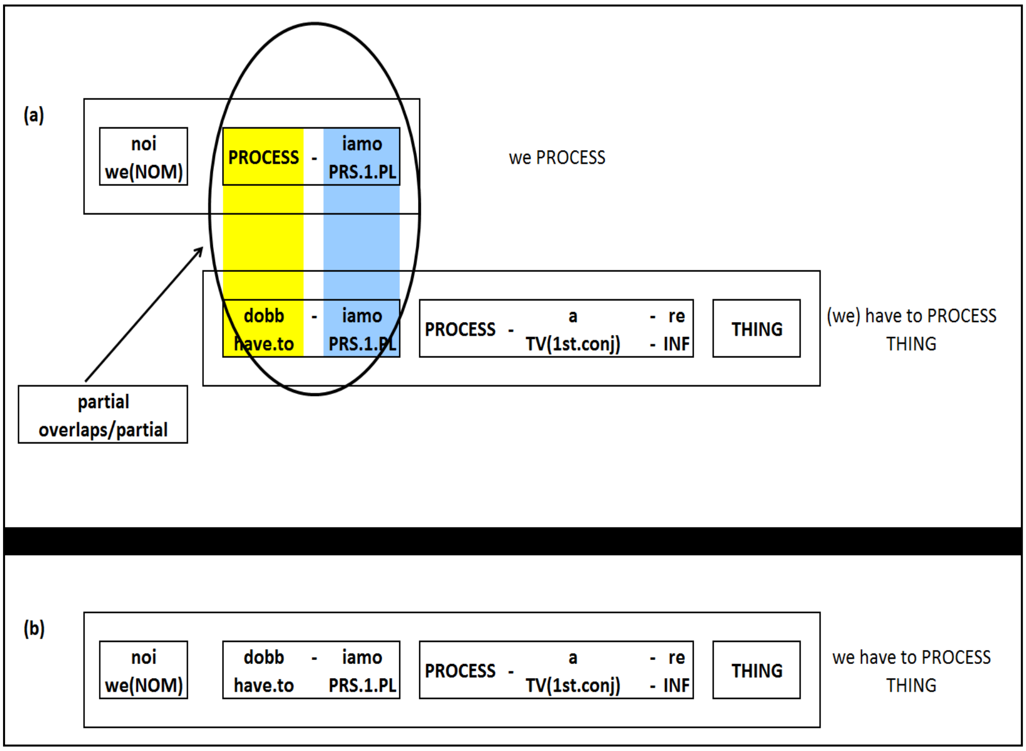


Figure 8.29: the partial overlap of 8.28b and 8.28c (section a) and the resulting string (section b). Slot elaboration is highlighted in yellow. Shared lexical material is highlighted in blue.

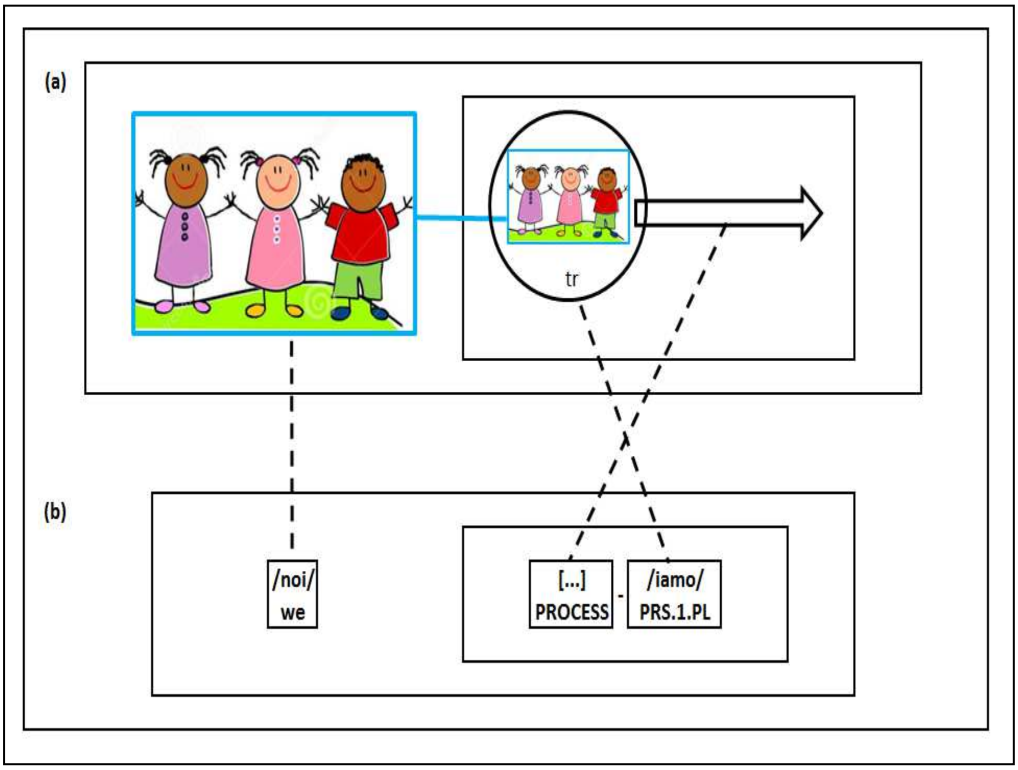


Figure 8.30: the semantic (a) and phonological pole (b) of the unit in fig. 8.28b. The dashed lines indicate relationships of symbolisation. At the semantic pole of both constituents (pronoun and verb), the agent (1.PL) is enclosed in a blue box. The two are linked by a blue line; this indicates that the pronoun *noi* "we" and the number-marker – *iamo* map onto the same meaning/entity. Translations into English are provided under phonetic transcriptions. [...] indicates that the unit is phonologically unspecified.

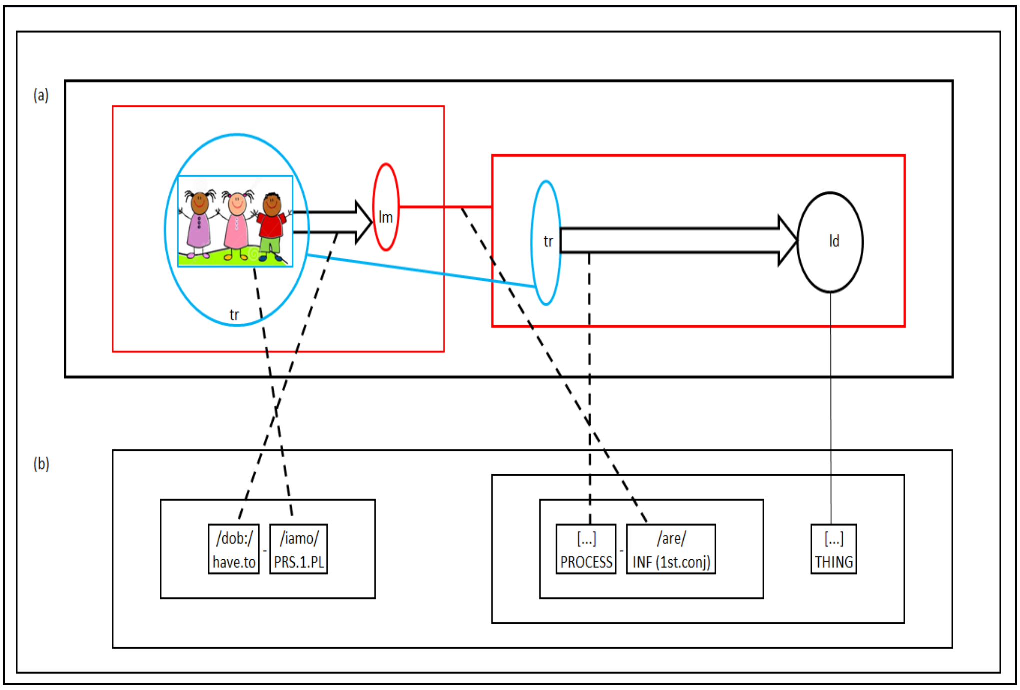


Figure 8.31: semantic (a) and phonological pole (b) of the unit in fig. 8.28c. Dashed lines indicate relationships of symbolisation. The red line that links the landmark of the auxiliary verb (in the red circle) and the dependent *PROCESS* (in the red rectangle) indicates that the infinitive clause is the landmark of the auxiliary. The infinitive marker *–are* is linked by symbolisation (dashed line) to the red line; this indicates that the infinitive marker maps onto the fact that the infinitive clause depends on the auxiliary. The blue box, circle and line indicate that the trajector (agent) of the auxiliary corresponds to the trajector of the dependent clause. Translations into English are provided under phonetic transcriptions. [...] indicates that the unit is phonologically unspecified.

70. *CHI: *stavo quasi cadendo nel buco.*

st-av-o	*quasi*	*cad-e-ndo*
stay-IMPERF-1.SG	almost	fall-TV(conj.II)-ing
ne=l	*buc-o*	
in=the(M.SG)	hole-M.SG	

"I almost fell into the hole."

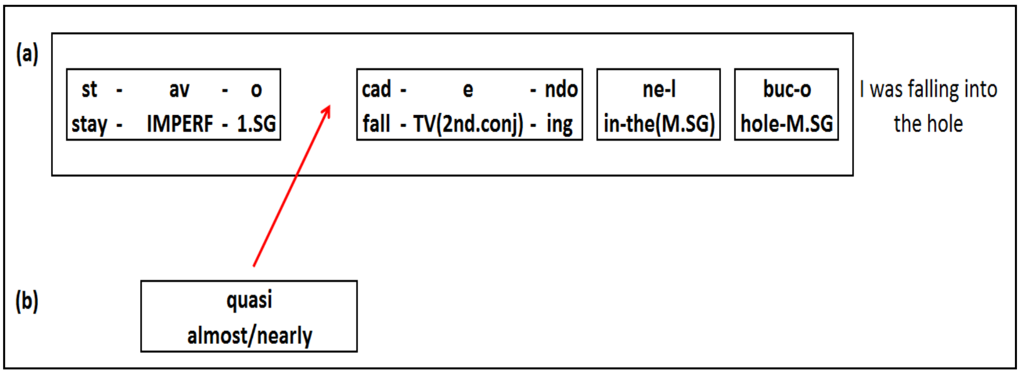


Figure 8.32: deriving target sentences SB128 through insertion.

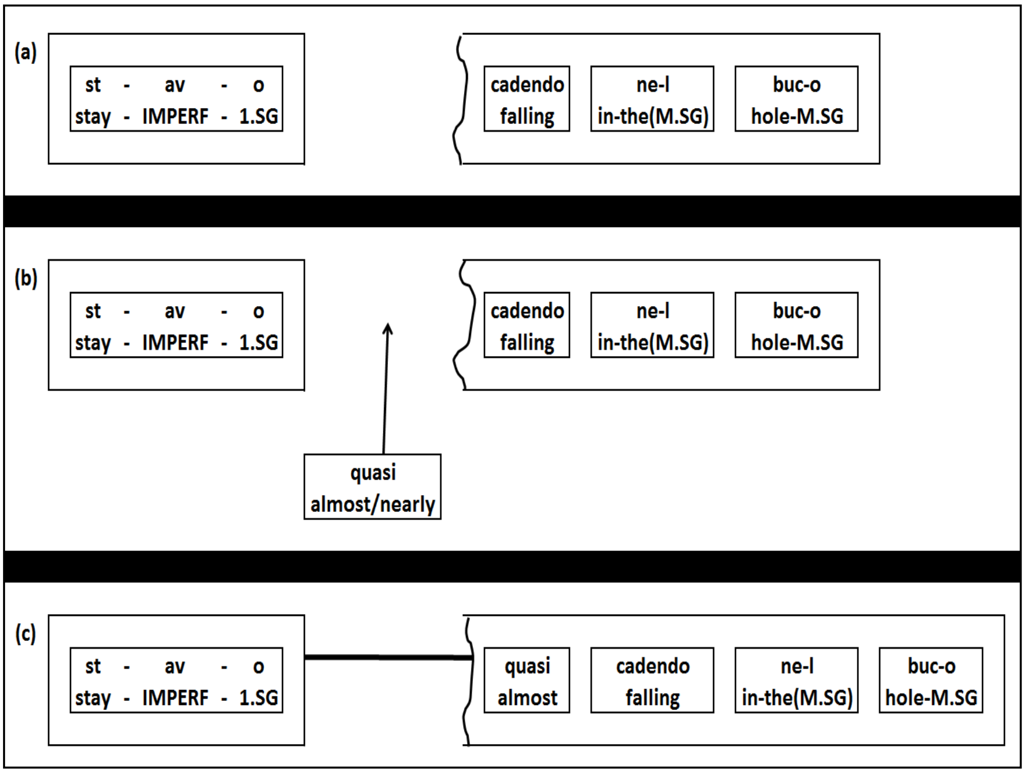


Figure 8.33: the operation of insertion (break, a; insert, b; and re-arrange, c).

71. **CHI: < sai> [/] sai che io (.) ho fatto una corsa (..) grande + e sono arrivato (.) dal Luca +. +. e ho fatto pf@o.*
sai che io ho fatto una corsa grande e sono arrivato dal Luca e ho fatto puf.
sa-i che io ho fatto un-a
know-PRS.2.SG that I have(PRS.1.SG) done a-F.SG
cors-a grand-e e
run-F.SG big-SG and
sono arriv-a-t-o da=l
be(PRS.1.SG) arrive-TV(conj.I)-PTCP-M.SG at=the(M.SG)
Luca e ho fatto puff.
Luca and have(PRS.1.SG) done puff.
("You know that I have done a big run and I have arrived at the Luca and I've done 'puff'?").
"Did you know that I ran a long way and got to Luca and went 'puff'?"
(week5.2014.02.18.B: lines 1423 and 1425)

72. **CHI: No però sei troppo piccola per entrare nella mia casetta*
no, però sei troppo piccol-a per
no , but be(PRS.2.SG) too small-F.SG to
entr-a-re ne=ll-a mi-a
enter-TV(conj.I)-INF in=the-F.SG my-F.SG
cas-ett-a
house-little(ENDR)-F.SG
"No, but you are too small to enter into my little house."
(week5.2014.2.18.B: lines 2334-2335)

Table 8.5: comparing the complexity of the t-units identified in successful derivations and constructional fails.

T-UNITS COMPLEXITY: comparing successful derivations and constructional fails			
measure of complexity	successful derivations mean (SE)	constructional fails mean (SE)	Wilcoxon rank sum test with continuity correction (one-tailed)
number of words per t-unit	4.29 (0.73)	*6.15 (0.23)*	*W=38***
number of finite clauses per t-unit	1.03 (0.01)	1.11 (0.04)	W=27**

** p<.05

19.3. Part III: AN EXPERIMENTAL INSIGHT INTO THE DEVELOPMENT OF MORPHO-SYNTACTIC COMPETENCE

19.3.1. METHOD

Table 11.1: the participants who took part in the experiment divided by age group.

age group	gender of participants		total no. of participants	age of participants (years:months:days)	
	males	females		mean	range
two-year-olds	9	7	16	02;08	02;02.23 - 03;00.25
three-year-olds	11	5	16	03;08	03;01.16 - 03;11.13
four-year-olds	9	12	21	04;05	04;00.28 - 05;00.00
adults	5	9	14	29;00	24;10.19 - 43;10.07
total	34	33	67		

1. *Ti facccio vedere una cosa che si chiama verbare.*

   "I'll show you something that is called *verb-are* [verb-INF]."

2. *Prova a dire verbare.*

   "Try to say *verb-are* [verb-INF]."

3. *Ti faccio vedere come si fa a verbare.*

   "I'll show you how to *verb-are* [verb-INF]."

4. *Guarda cosa la [AGENT'S NAME] fa alla [PATIENT'S NAME]!*

   "look at what [AGENT'S NAME] is going to to [PATIENT'S NAME]!"

5. *Hai visto cosa la [AGENT'S NAME] ha fatto alla [PATIENT'S NAME]?*

   "Have you seen what [AGENT'S NAME] did to [PATIENT'S NAME]?"

6. V, A

Verba, [AGENT'S NAME]!
verb-a [AGENT'S NAME]!
verb-IMP.2.SG [AGENT'S NAME]!
"Verb, [AGENT'S NAME]!"

7. Vo

Verbala!
verb-a=l-a!
verb-IMP.2.SG=clitic.3.ACC-F.SG!
"Verb her!"

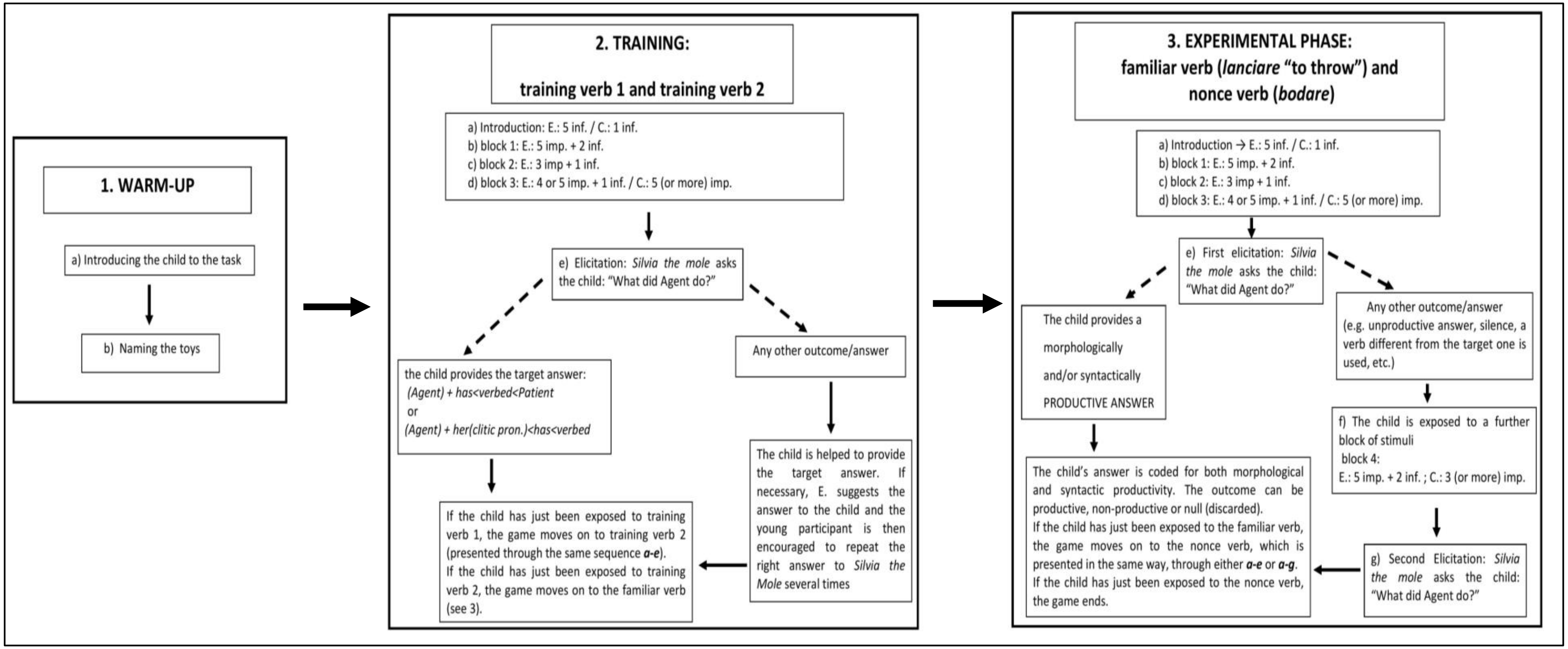


Figure 11.1: the design of the experimental study. Solid arrows indicate the order of the various sections which participants went through. Dashed arrows indicate sequences as consequences of binary possibilities. Hence, after the elicitation trial (2e and 3e), the game would move on differently, depending on the answer (or non-answer) provided by the child. In sections 2 and 3, the letters E. and C. refer to the Experimenter and the child, respectively. "Imp." indicates an imperative stimulus, presented within either a "Vo" or a "VA" construction. "Inf." indicates an infinitive form of the target verb. Hence, in "d) block 3" of both (2) training and (3) experimental phase, the sequence "E.: 4 or 5 imp. + inf./ C.: 5 or more imp." indicates that the Experimenter uttered 4 or 5 imperative forms and one infinitive form of the target verb and that the child uttered 5 or more imperative forms of the same verb. For what is classified as a productive (or non-productive) answer, see the coding section (11.4). For a more detailed illustration of the task, see chapter 21.3 (Appendix_III, Volume II).

8. a) “Vo” stimulus

lanciala!
lanci-a=l-a!
throw-IMP.2.SG=clitic.3.ACC-F.SG
“Throw her!”

b) “V,A” stimulus
lancia, Peppa
lanci-a, *Peppa*
throw-IMP.2.SG, Peppa
“Throw, Peppa!”

9. a) “(S)+AUX<V<O” output

(Peppa) *ha lanciato (la) Emily*
(Peppa) *ha* *lanci-a-to* *(l-a)* *Emily*
(Peppa) has throw-TV(conj.I)-PTCP (the-F.SG) Emily
“(Peppa) has thrown Emily.”

b) “(S)+<o<AUX<V” output
(Peppa) l’ ha lanciata
(Peppa) *l’* *ha* *lanciat-a-t-a*
(Peppa clitic.3.ACC has throw-TV(conj.I)-PTCP-F.SG
“(Peppa) has thrown her.”

Table 11.2: participants and the experimental conditions to which they contributed.

age group	gender		no of participants in each age group	participants' contribution			no of answers for the familiar verb	no of answers for the nonce verb
	m	f		familiar and nonce verb	familiar verb only	nonce verb only		
two-year-olds	9	7	16	9	4	3	13	12
three-year-olds	11	5	16	15	1	0	16	15
four-year-olds	9	12	21	15	1	5	16	20
adults	5	9	14	14	0	0	14	14

Table 11.3: the distribution of construction conditions across verb_familiarity and age_group.

age_group	familiar verb		nonce verb	
	"Vo"	"V,A"	"Vo"	"V,A"
two-year-olds	7	6	4	8
three-year-olds	6	10	10	5
four-year-olds	9	7	9	11
adults	7	7	7	7

Table 11.4: the distribution of agent and construction conditions across verb_familiarity and age_group.

verb_familiarity	familiar verb						nonce verb					
agent	Peppa_Pig		Emily_Elefante		chosen by the participant		Peppa_Pig		Emily_Elefante		chosen by the participant	
construction	"Vo"	"V,A"	"Vo"	"V,A"	"Vo"	"V,A"	"Vo"	"V,A"	"Vo"	"V,A"	"Vo"	"V,A"
two-year-olds	3	4	0	1	4	1	2	5	0	1	2	2
three-year-olds	2	6	4	3	0	1	6	2	3	2	1	1
four-year-olds	5	3	4	3	0	1	4	4	4	7	1	0
adults	3	4	4	3	0	0	3	4	4	3	0	0

19.3.2. RESULTS

Table 13.1: Results by age group. As discussed in chapter 11, some children contributed data for only one or the other verb_familiarity condition, hence the unequal number of answers between nonce and familiar verbs (refer back to table 11.2).

morphology						
age group	familiar verb			nonce verb		
	prod	not prod	tot	prod	not prod	tot
two-year-olds	8	5	13	4	8	12
three-year-olds	15	1	16	11	4	15
four-year-olds	16	0	16	16	4	20
adults	13	1	14	14	0	14
syntax						
age group	familiar verb			nonce verb		
	prod	not prod	tot	prod	not prod	tot
two-year-olds	8	5	13	2	10	12
three-year-olds	14	2	16	6	9	15
four-year-olds	16	0	16	14	6	20
adults	13	1	14	13	1	14

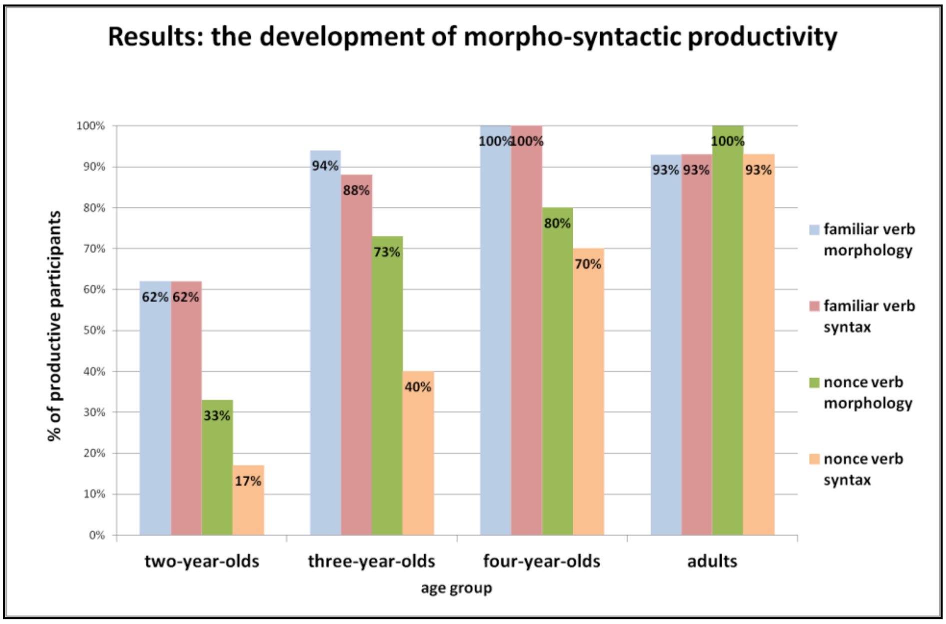


Figure 13.1: morphological and syntactic productivity: summary of results.

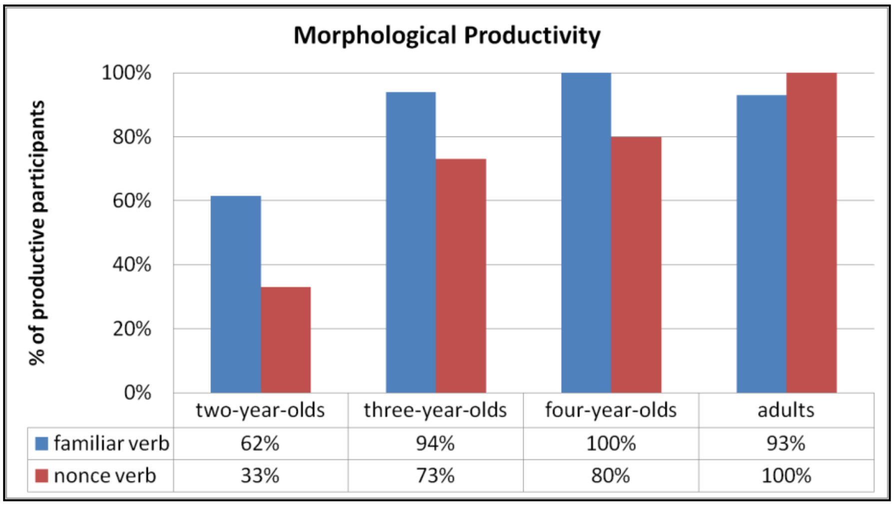


Figure 13.2: morphological productivity as a function of age_group and verb_familiarity.

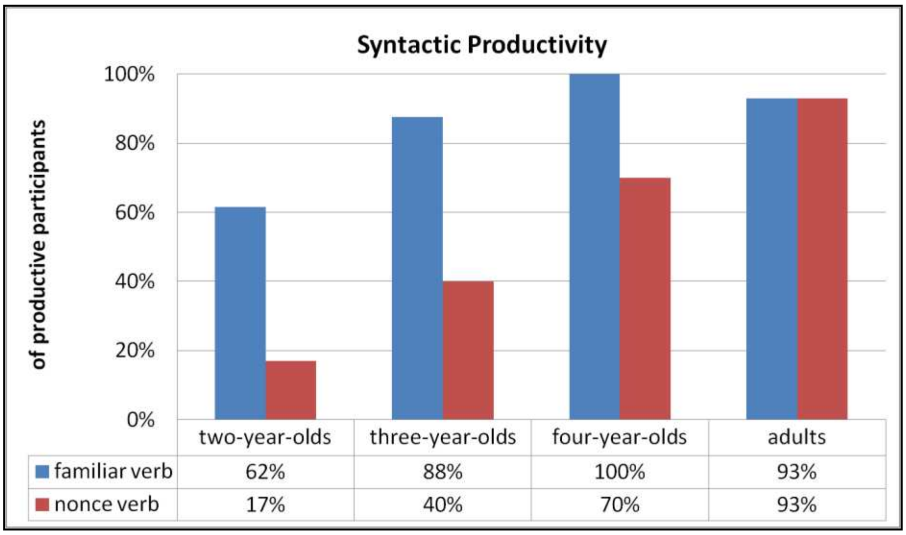


Figure 13.3: syntactic productivity as a function of age_group and verb_familiarity.

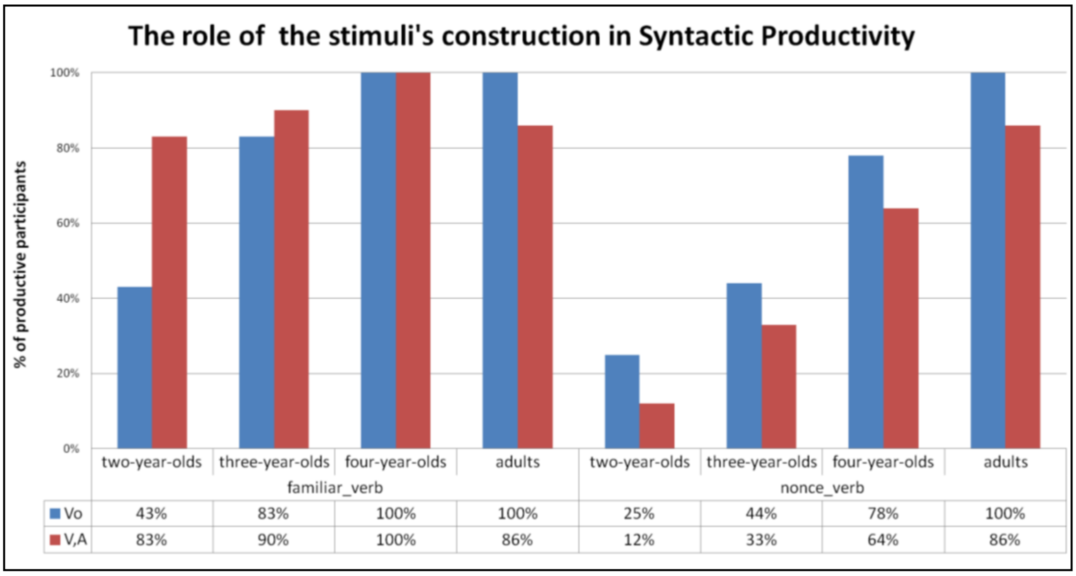


Figure 13.4: syntactic productivity as a function of age_group, verb_familiarity and construction.

19.3.3. ANALYSIS

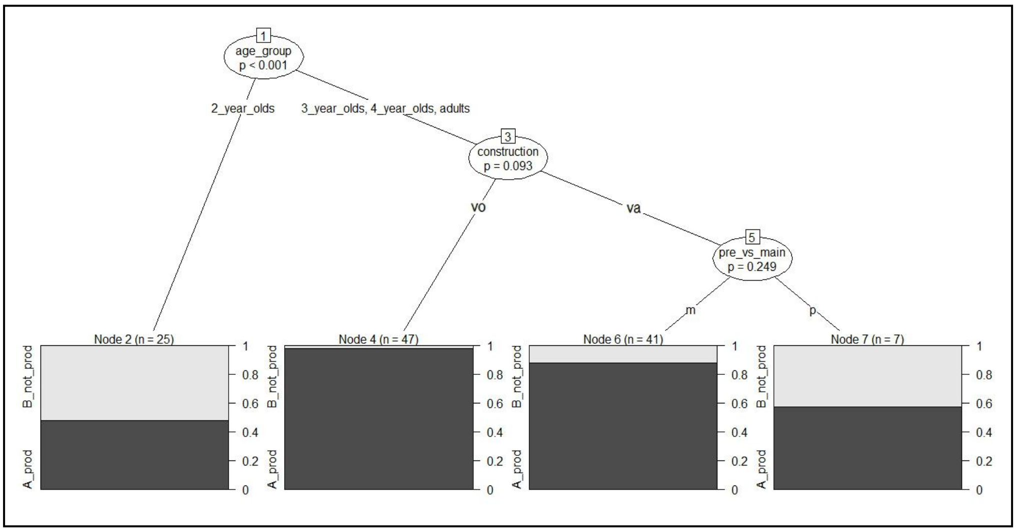


Figure 14.1: Results pertaining to morphological productivity.

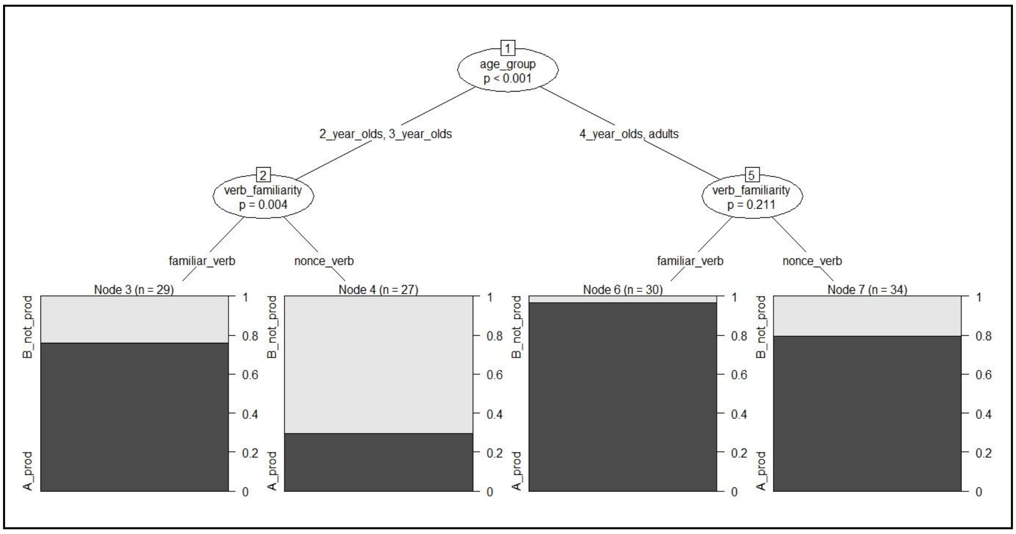


Figure 14.2: Results pertaining to syntactic productivity.

Table 14.1: Mixed-effects logistic regressions. For each model, productivity is the dichotomous (productive vs. not_productive) DV and verb_familiarity the dichotomous predictor (familiar vs nonce). The only random effect is participants.

MIXED-EFFECT MODELS: productiviy ~ verb_familiarity + (1|participants)				
(a) MORPHOLOGICAL PRODUCTIVITY				
age_group	Intercept Estimate (Std. Error)	Intercept z-value	verb_familiarity (familiar vs nonce) Estimate (Std. Error)	verb_familiarity (familiar vs nonce) z-value
two-year-olds	2.07 (2.94)	0.70	-4.68 (4.54)	-1.03
three-year-olds	**2.71 (1.03)**	**2.62****	-1.32 (1.22)	-1.09
four-year-olds	22.95 (648.86)	0.03	-21.51 (639.30)	-0.03
adults	2.565e+00 (1.038e+00)	2.47	3.459e+01 (1.794e+07)	0
(b) SYNTACTIC PRODUCTIVITY				
age_group	Intercept Estimate (Std. Error)	Intercept z-value	verb_familiarity (familiar vs nonce) Estimate (Std. Error)	verb_familiarity (familiar vs nonce) z-value
two-year-olds	0.47 (0.57)	0.82	***-2.08 (0.96)***	***-2.16*****
three-year-olds	**15.23 (6.97)**	**2.16****	***-18.17 (8.22)***	***-2.21*****
four-year-olds	30.31 (2053)	0.015	-26.44 (2063)	-0.014
adults	**2.565e+00 (1.038e+00)**	**2.47****	-2.387e-15 (-2.387e-15)	0

*p<.1; **p<.05

Table 14.2: Pair-wise comparisons across age groups with the familiar verb (left) and the nonce verb (right) with respect to morphological (upper tables) and syntactic (bottom tables) productivity. P-values are adjusted with Bonferroni corrections (*p<.1 and **p<.05).

MORPHOLGY - FAMILIAR VERB: Fisher's Tests	adults	four-year-olds	three-year-olds	two-year-olds
adults				
four-year-olds	0.467 (0 - 34.13)			
three-year-olds	0.87 (0.01 - 73.16)	Inf (0.03 - Inf)		
two-year-olds	7.52 (0.68 - 412.02)	Inf (1.38 - Inf)*	8.66 (0.78 - 470 .74)	

MORPHOLGY - NONCE VERB: Fisher's Tests	adults	four-year-olds	three-year-olds	two-year-olds
adults				
four-year-olds	Inf (0.48 - Inf)			
three-year-olds	Inf (0.67 - Inf)	1.09 (0.22 - 9.59)		
two-year-olds	**inf (3.62 - Inf) ****	7.39 (1.25 - 55.41)	5.12 (0.82- 39.75)	

SYNTAX - FAMILIAR VERB: Fisher's Tests	adults	four-year-olds	three-year-olds	two-year-olds
adults				
four-year-olds	0 (0 - 34.13)			
three-year-olds	1.82 (0.09 - 117.77)	Inf (0.19 - Inf)		
two-year-olds	7.52 (0.68 - 412.02)	Inf (1.38 - Inf)*	4.14 (0.53 - 53.21)	

SYNTAX - NONCE VERB: Fisher's Tests	adults	four-year-olds	three-year-olds	two-year-olds
adults				
four-year-olds	5.33 (0.53 - 275.15)			
three-year-olds	**17.4 (1.76 - 912.82) ****	3.37 (0.70 - 18.03)		
two-year-olds	**48.73 (4.03 - 294.60)****	10.66 (1.59 - 129.40)*	3.19 (0.42 - 40.28)	

*p < .1 **p < .05

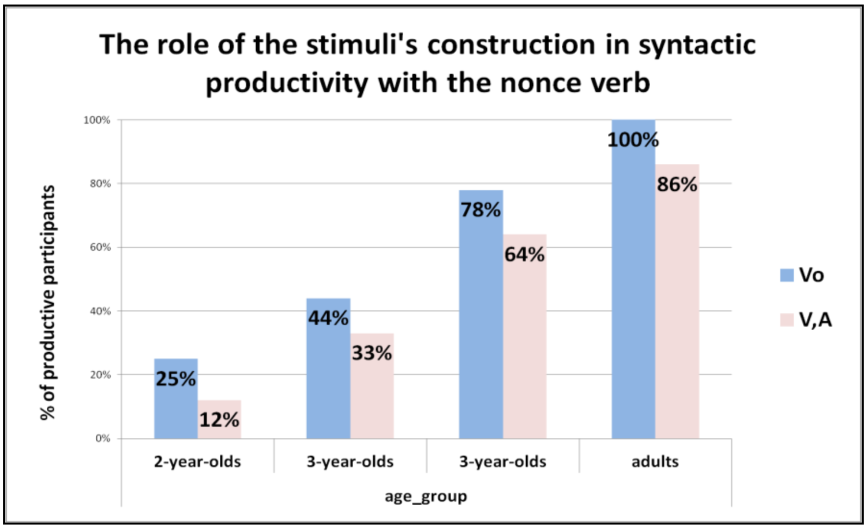


Figure 14.3: proportion of syntactically productive participants with the nonce verb depending on the construction with which stimuli were presented.

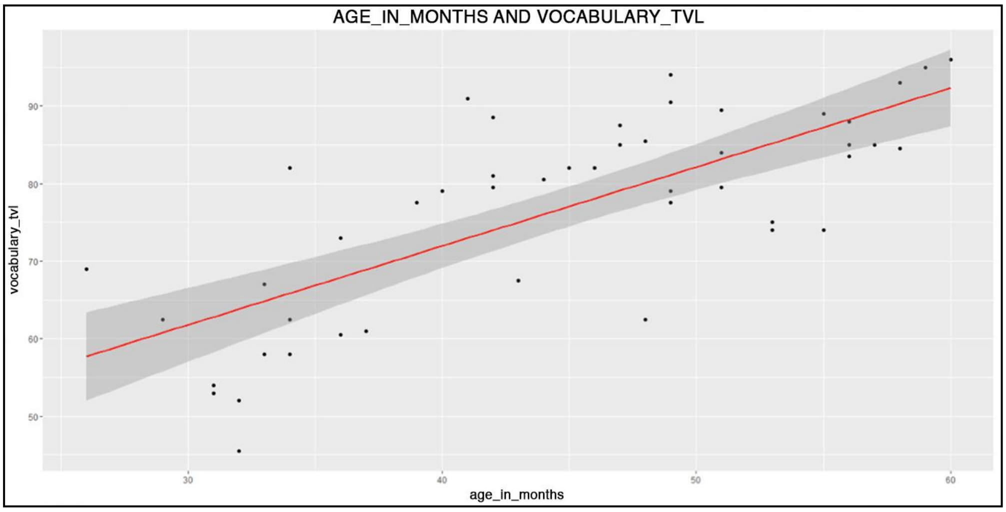


Figure 14.4 The relationships between age_in_months (x-axis) and vocabulary_tvl (y-axis).

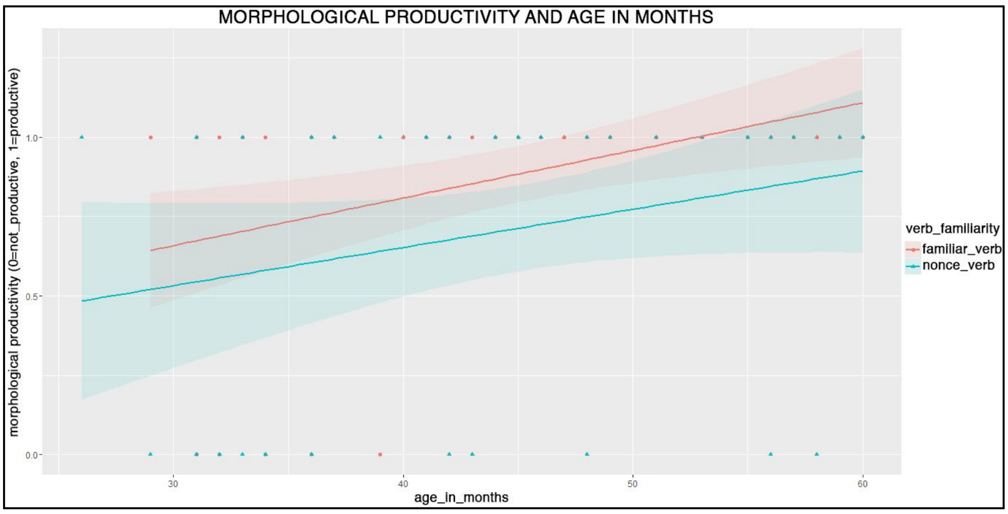


Figure 14.5 : morphological productivity with familiar and nonce verbs as a function of age_in_months.

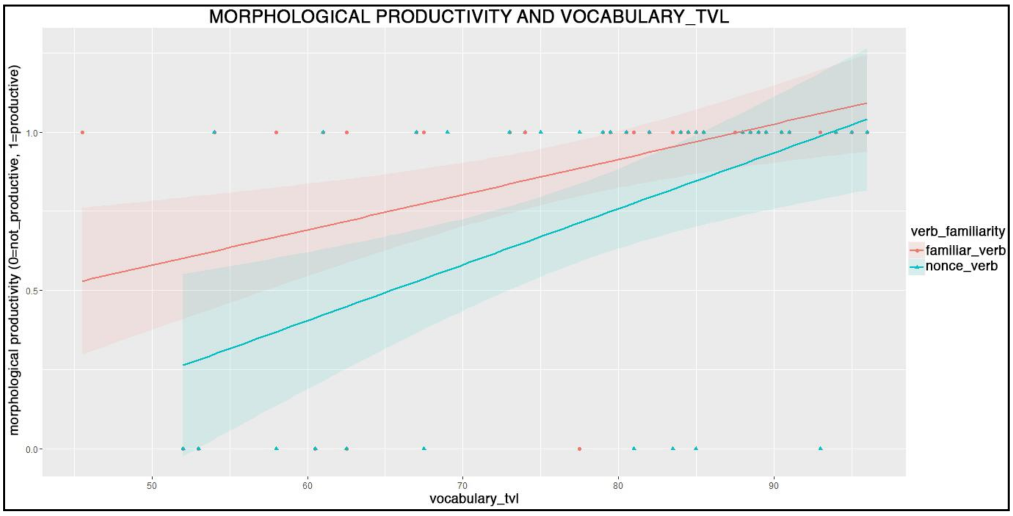


Figure 14.6: morphological productivity with familiar and nonce verbs as a function of vocabulary_tvl.

Table 14.3: Robust t-tests that compare vocabulary_tvl and age_in_months between morphologically productive and morphologically non-productive children. For each subtable (age in months and vocabulary score), Column 1 (lines 2 and 3) indicates verb familiarity. Column 2 reports the number (N=) of children who were productive, the 20% trimmed mean of their age in months (top subtable) and their vocabulary scores (bottom subtable). Column 3 reports the number (N=) of children who were not productive, the 20% trimmed mean of their age in months (top subtable) and their vocabulary scores (bottom subtable). Column 4 reports the difference in trimmed means between productive and non-productive children (and their CIs) regarding the production with familiar verb (line 2) and nonce verb (line 3).

MORPHOLOGICAL PRODUCTIVITY: comparing productive and non-productive children (t-tests)			
age in months			
verb_familiarity	productive participants: N (20%) trimmed mean	non-produtive participants: N (20%) trimmed mean	difference between (20%) trimmed means (CIs)
familiar verb	***N=38***	***N=5***	***13.13****
	47	***43***	***(7.88 - 17.13)***
nonce verb	N=30	N=12	6.72
	47.22	40.05	(-2.25 - +14.13)
vocabulary score (tvl)			
verb_familiarity	productive participants: N (20%) trimmed mean	non-produtive participants: N (20%) trimmed mean	difference between (20%) trimmed means (CIs)
familiar verb	***N=38***	***N=5***	***22.94****
	81.60	***58.67***	***(8.83 - 29.73)***
nonce verb	***N=30***	***N=12***	***15.03****
	88.28	***67.25***	***(5.25 - 23.47)***

* p<.05

Table 14.4: point bi-serial correlations and partial point bi-serial correlations between morphological productivity and either age_in_months or vocabulary_tvl. Results significant at p < .05 are *in bold and Italics*. In each sub-table (a and b), the relationship between productivity and either age_in_months (left) or vocabulary_tvl (right) is reported. In each sub-table, it is reported whether results pertain to productivity with the familiar verb (line 2) or the nonce verb (line 3). Columns 2 and 6 report the test statistics (r_{pb}), columns 3 and 7 report the R-squared of the relationship, columns 4 and 8 report the % of variance explained by either age_in_months or vocabulary_tvl and columns 5 and 9 report the p-value of the test statistics.

MORPHOLOGICAL PRODUCTIVITY (correlation analyses)								
(a) correlations								
verb_familiarity	age_in_months				vocabulary_tvl			
	r_{pb}	R-squared	explained variance	p-value	r_{pb}	R-squared	explained variance	p-value
familiar verb	**.425**	**.180**	**18%**	**.005**	***.452***	***.204***	***20%***	***.002***
nonce verb	.247	.061	6%	.115	***.474***	***.225***	***23%***	***.002***
(b) partial correlations								
verb_familiarity	age_in_months (vocabulary_tvl held constant)				vocabulary_tvl (age_in_months held constant)			
	r_{pb}	R-squared	explained variance	p-value	r_{pb}	R-squared	explained variance	p-value
familiar verb	.143	.020	2%	=.367	.221	.049	5%	.160
nonce verb	-.221	.045	5%	=.185	***.460***	***.211***	***21%***	***.002***

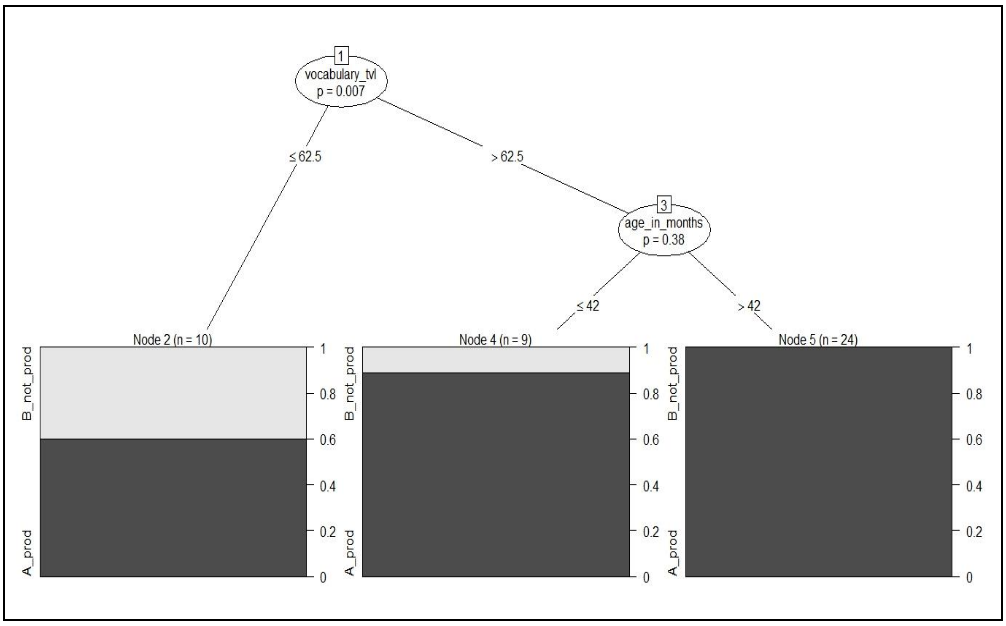


Figure 14.7: morphological productivity with the familiar verb as a function of vocabulary_tvl and age_in_months.

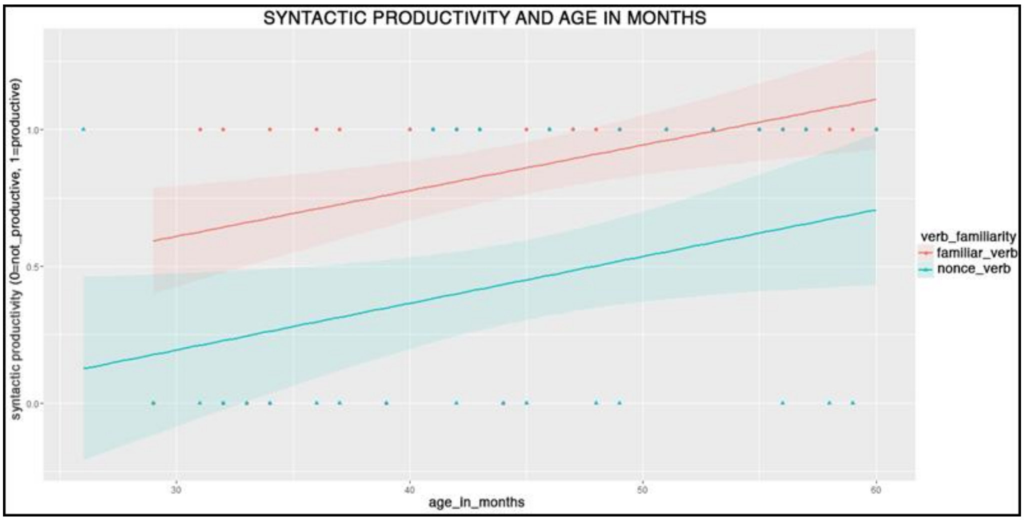


Figure 14.8: syntactic productivity with familiar and nonce verbs as a function of age_in_months.

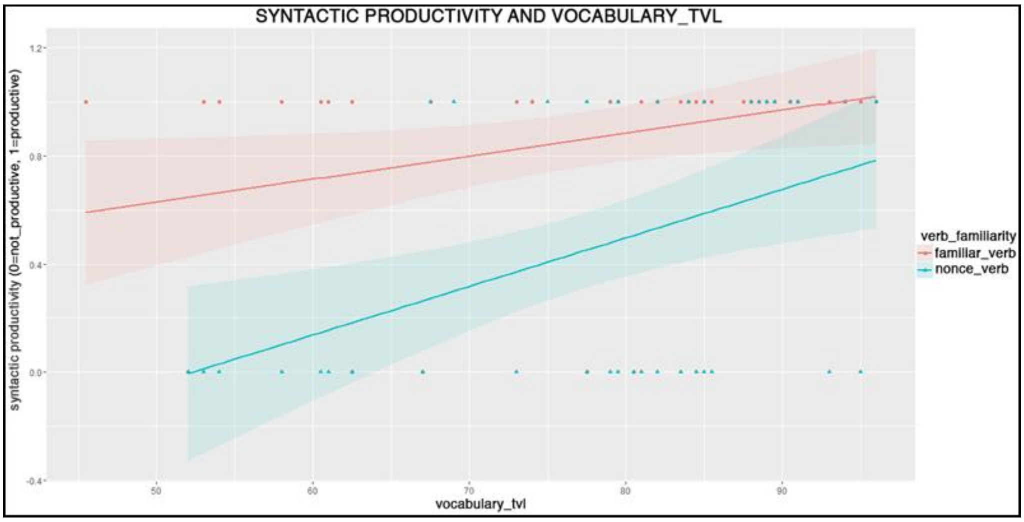


Figure 14.9: syntactic productivity with familiar and nonce verbs as a function of vocabulary_tvl.

Table 14.5: Robust t-tests that compare vocabulary_tvl and age_in_months between syntactically productive and syntactically non-productive children. For each subtable (age in months and vocabulary score), Column 1 (lines 2 and 3) indicates verb familiarity. Column 2 reports the number (N=) of children who were productive, the 20% trimmed mean of their age in months (top subtable) and their vocabulary scores (bottom subtable). Column 3 reports the number (N=) of children who were not productive, the 20% trimmed mean of their age in months (top subtable) and their vocabulary scores (bottom subtable). Column 4 reports the difference in trimmed means between productive and non-productive children (and their CIs) regarding production with familiar verb (line 2) and nonce verb (line 3).

SYNTACTIC PRODUCTIVITY: comparing productive and non-productive children (t-tests)			
age in months			
verb_familiarity	productive participants: N (20%) trimmed mean	non-produtive participants: N (20%) trimmed mean	difference between (20%) trimmed means (CIs)
familiar verb	***N=37***	***N=6***	***12.76****
	47.26	***34.5***	***(6.18 - 17.48)***
nonce verb	***N=19***	***N=23***	***7.69****
	49.16	***41.47***	***(0.90 - 12.90)***
vocabulary score (tvl)			
verb_familiarity	productive participants: N (20%) trimmed mean	non-produtive participants: N (20%) trimmed mean	difference between (20%) trimmed means (CIs)
familiar verb	***N=37***	***N=6***	***14.28****
	81.65	***67.38***	***(3.68 - 23.93)***
nonce verb	**N=19**	**N=23**	**10.94***
	84.04	**73.10**	**(3.11- 18.41)**

* p<.05

Table 14.6: point bi-serial correlations and partial point bi-serial correlations between syntactic productivity and either age_in_months or vocabulary_tvl. Results significant at p < .05 are *in bold and Italics*. In each sub-table (a and b), the relationship between productivity and either age_in_months (left) or vocabulary_tvl (right) is reported. In each sub-table, it is reported whether results pertain to productivity with the familiar verb (line 2) or the nonce verb (line 3). Columns 2 and 6 report the test statistics (r_{pb}), columns 3 and 7 report the R-squared of the relationship, columns 4 and 8 report the % of variance explained by either age_in_months or vocabulary_tvl and columns 5 and 9 report the p-value of the test statistics.

SYNTACTIC PRODUCTIVITY (correlation analyses)								
(a) correlations								
verb_familiarity	age_in_months				vocabulary_tvl			
	r_{pb}	R-squared	explained variance	p-value	r_{pb}	R-squared	explained variance	p-value
familiar verb	***.437***	***.191***	***19%***	***.003***	***.319***	***.102***	***10%***	***.037***
nonce verb	***.319***	***.101***	***10%***	***.040***	***.437***	***.191***	***19%***	***.004***
(b) partial correlations								
verb_familiarity	age in months (vocabulary_tvl held constant)				vocabulary_tvl (age_in_months held constant)			
	r_{pb}	R-squared	explained variance	p-value	r_{pb}	R-squared	explained variance	p-value
familiar verb	***.317***	***0.100***	***10%***	***.041***	-.020	-.000	0%	.899
nonce verb	-.032	-.001	0%	.842	***.318***	***.101***	***10%***	***.043***

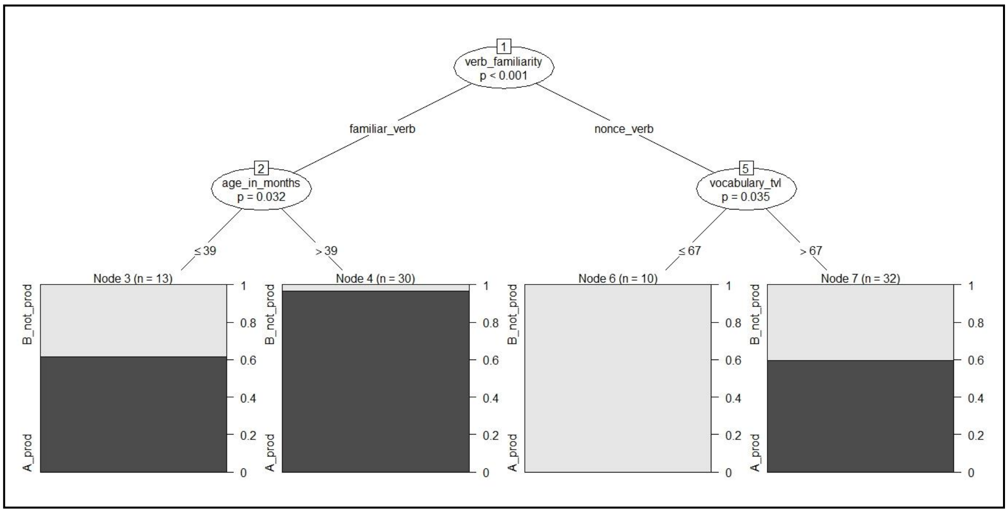


Figure 14.10 syntactic productivity as a function of vocabulary_score, age_in_months, verb_familiarity, construction, pre_vs_main, school and gender, when only children whose vocabulary data are available are considered.

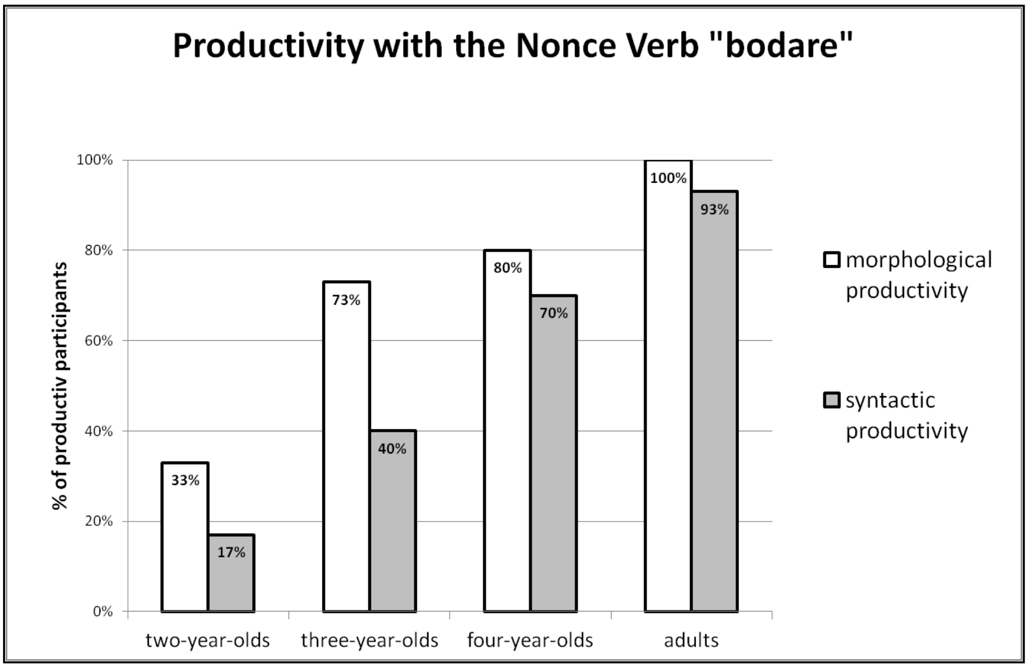


Figure 16.1: The proportion of productive participants with the nonce verb in each age-group.

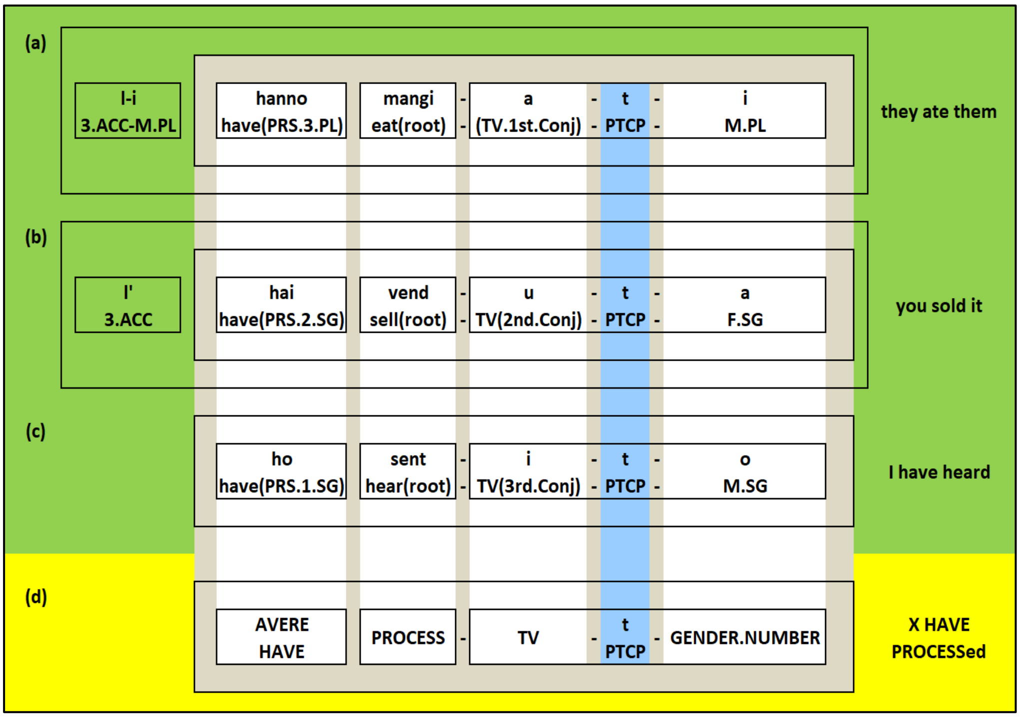


Figure 16.2: Abstracting the *passato prossimo* schema (in the grey square). Concrete strings are included in the green strip, whereas the schema appears in the yellow strip. The blue highlighting indicates fixed recurring material that consistently maps onto a fixed meaning (past). Slot formation is highlighted in white.

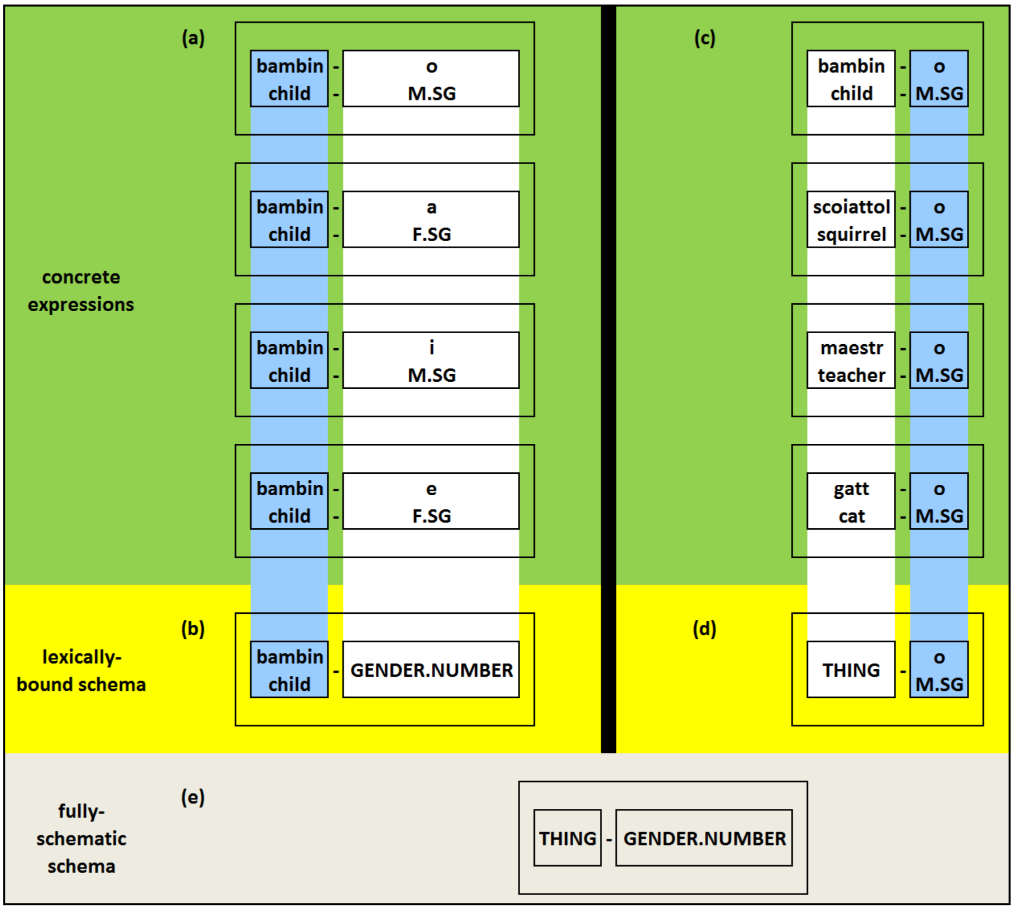


Figure 16.3: abstracting morphological schemas in Italian. Concrete expressions are in the green strip and lexically-bound schemas in the yellow strip. The fully schematic unit abstracted from the two lexically-specific schemas is in the grey strip (e). Shared lexical material is in blue and relationships of elaboration (i.e. slot formation) are highlighted in white.

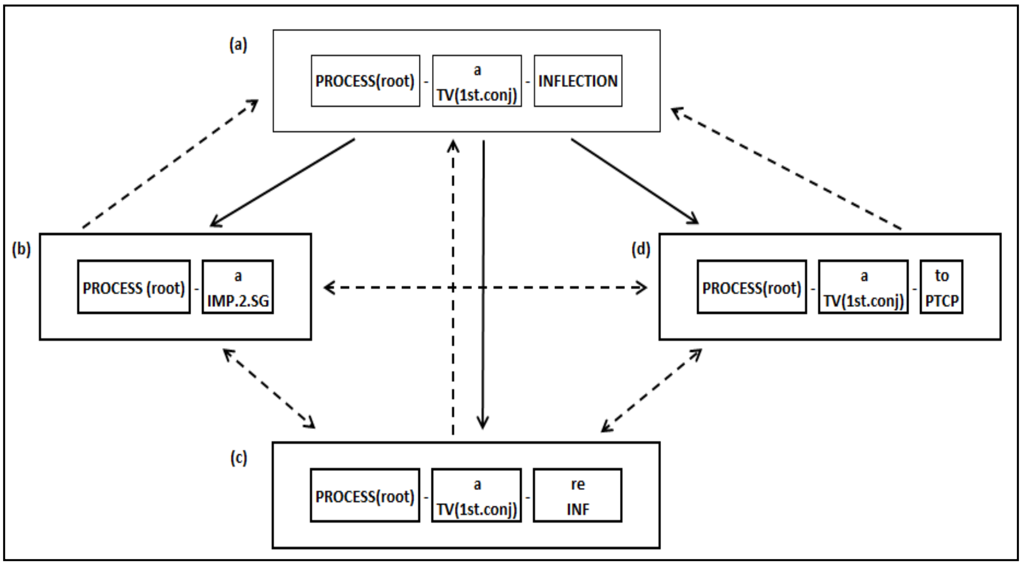


Figure 16.4: the putative network of morphological constructions that could be hypothesised as being developed by two-year-old Italian-speaking children. Dashed arrows indicate relationships of extension. Solid arrows indicate relationships of elaboration. Thickness of boxes indicates units' entrenchment.

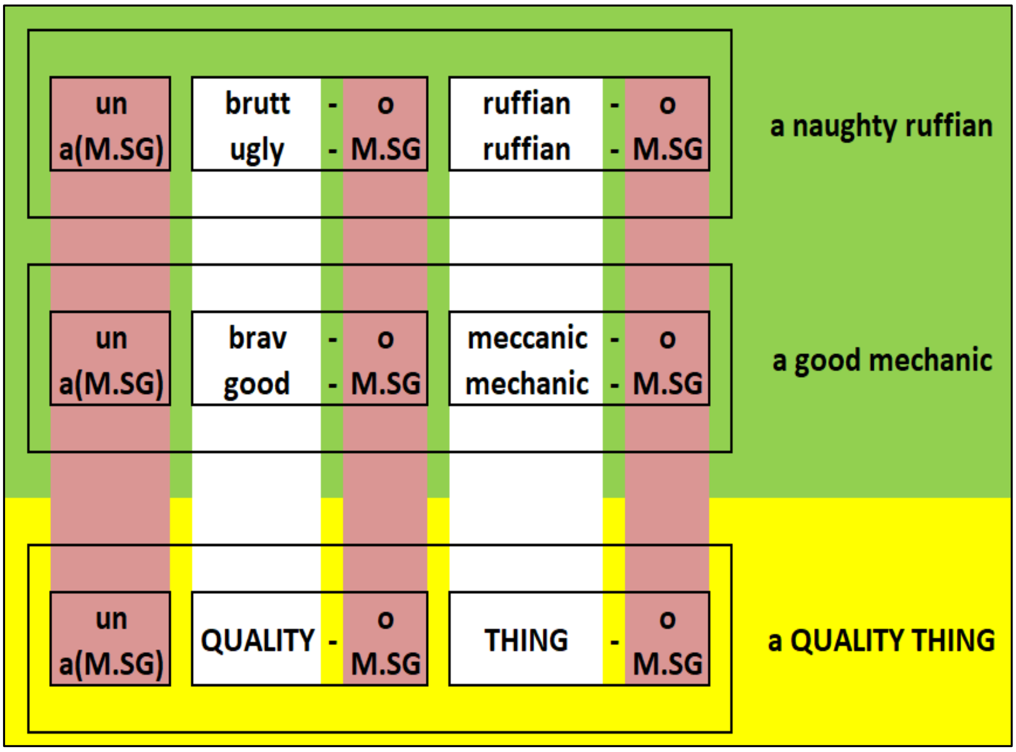


Figure 16.5: an attested schema that accounts for gender-number agreement across article, adjective and noun. Slot schematisation is in white. The concrete and recurring material that accounts for gender-number agreement is highlighted in red.

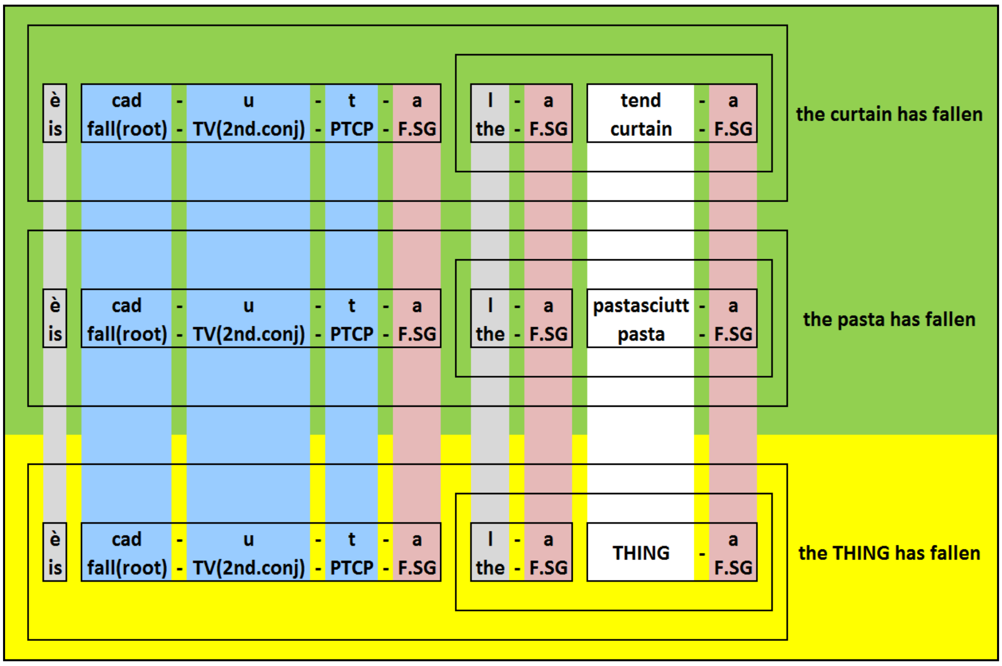


Figure 16.6: an attested schema that accounts for gender-number agreement across verb, past participle and subject-NP. Concrete and recurring lexical material which is not essential for agreement is in blue. Concrete and recurrent material that accounts for number agreement only is highlighted in grey. Finally, the red highlighting indicates concrete and recurring material that accounts for both gender and number agreement across elements. Slot schematisation is highlighted in white.

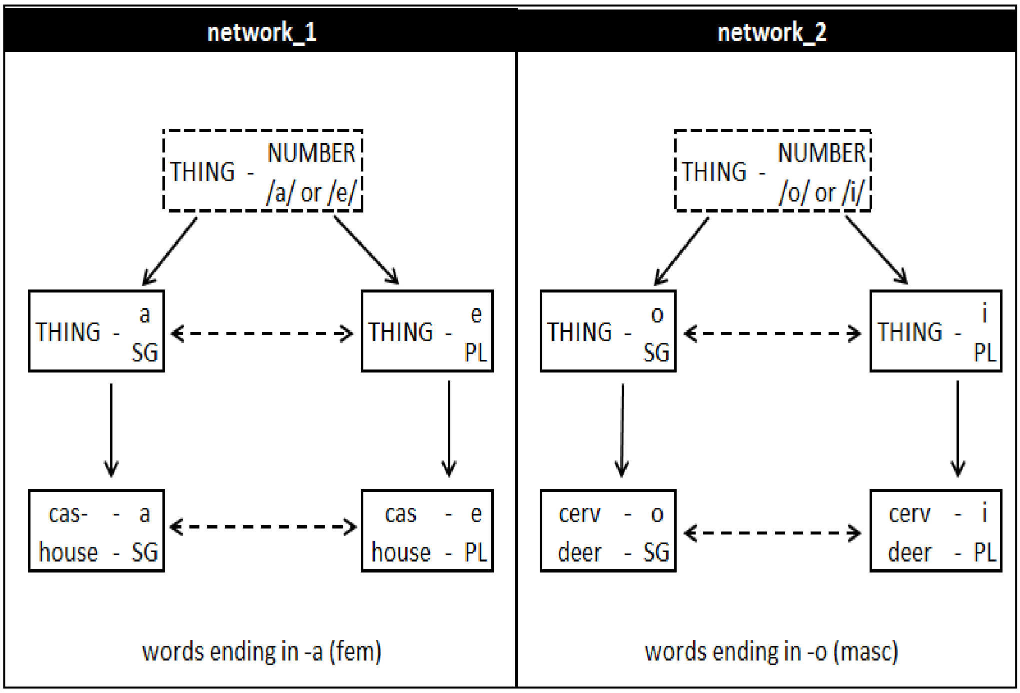


Figure 16.7: two independent networks of morphological constructions that account for gender-number flexibility. Network 1 can be paraphrased as "words that end in *–a* take plural form *–e*". Network_2 can be paraphrased as "words that end in *–o* take plural form in *–i*". Relationships of elaboration and extension are indicated by solid and dashed arrows, respectively. Thickness of boxes indicates units' entrenchment.

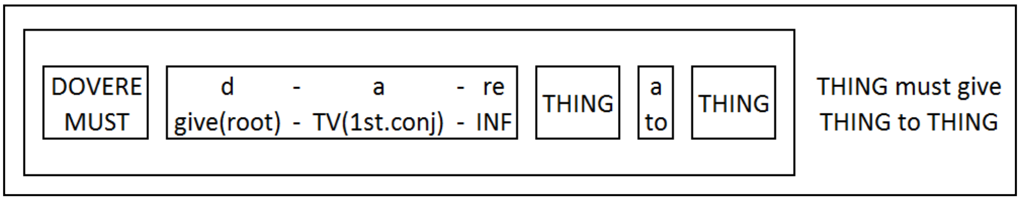


Figure 16.8: A three-slot schema that could account for one of the constructional fails.

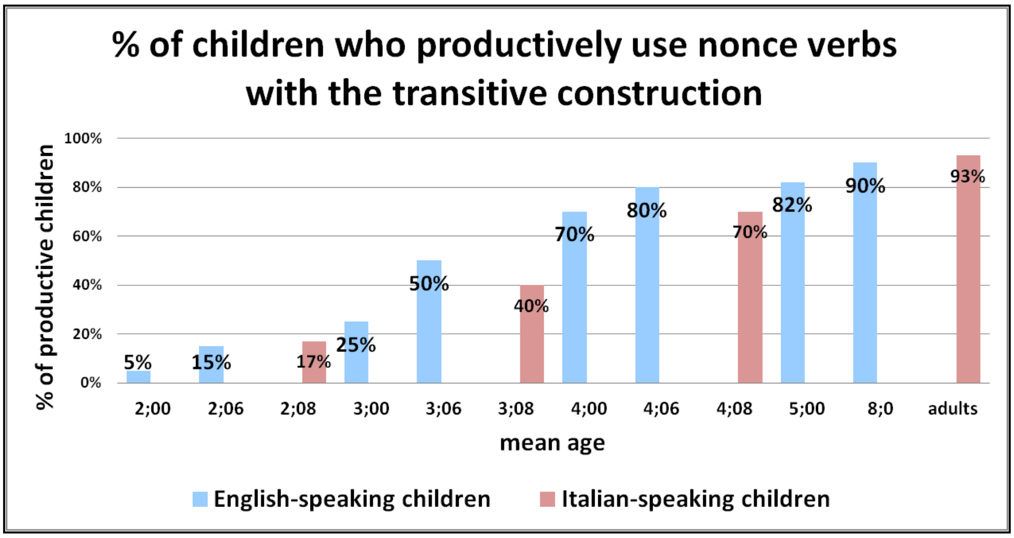


Figure 16.9: proportion of English-speaking children (blue; adapted from Tomasello, 2006b, fig. 6.1, p. 266) and Italian-speaking children (red; refer to fig. 13.3) who produce grammatical transitive utterances with nonce verbs.

Page intentionally left blank

20.

APPENDIX_II:

The spontaneous production of a two-year-old child

Page intentionally left blank

20.1. CHAPTER 6: METHOD

Sara (mother) and Sebastiano (father) addressed their son only in Italian throughout the whole study, with the exception of a few Trentin words, such as *vara* (Italian *guarda*: look-IMP.2.SG). In a few, extremely rare occasions, Sebastiano addressed his son with Trentin.

Throughout the length of the study, paternal grandparents overwhelmingly used Trentin when they spoke to each other and to Sebastiano (who in turn used Trentin with them), whereas they used Italian when addressing Roberto. Instances of code switching depending on to whom grandparents were talking are frequent in the recordings in which they appear, as the example below shows[4] (from the recording of 01.02.14):

450 *GRM: *prepariamo il tavolo, che dopo quando è pronto +...*
prepar-iamo il tavol-o che
Get.ready-PRS1.PL the(M.SG) table-M.SG that
dopo quando è pront-o +...
after when is ready-M.SG +…
"let's set the table, so later on, when it's ready +…"

451 *FAT: +^ *ma quello è il dolcificante.*
ma quell-o è il
but that-M.SG is the(M.SG)
dolcificant-e
sweetener(M.)-SG
"but that one is sweetener"

452 *FAT: *dai, sta fermo, tu!*
da-i, st-a ferm-o,
give-PRS.2.SG, stay-IMP.2.SG still-M.SG,
tu!
you(2.SG.NOM.)!
"C'mon, don't move, YOU!"

453 *GRM: fermo che quello lì è della nonna!
ferm-o che quell-o lì è
still-M.SG that that-M.SG there is
de(l)=l-a nonn-a
of=the-F.SG grandma-F.SG
"don't move! That one is Grandma's!"

454 *GRM: *mettilo via!*
mett-i=l-o via!
put.IMP.2.SG=clitic.3.ACC-M.SG away

[4] Code switching is indicated in brackets [- tre] and sentences in Trentin have also been emboldened.

"put it away!"

455 *GRM: ***[- tre] è perchè ghé digo che l'è<zucchero@s:ita> [>].***

è	*perchè*	*ghè*	*dig-o*
is	because	clitic.DAT.3	say-PRS.1SG
che	*l'*	*è*	*zuccher-o*
that	clitic.3.SG.NOM.	is	sugar-M.SG

"It's because I tell him that it's sugar"

On lines 450 and 453, the grandmother is speaking to Roberto and uses Italian. On line 455 she speaks to Sebastiano and switches to Trentin.

Although both grandparents mainly address Roberto with standard Northern Italian, they (very occasionally) also use Trentin with him.

20.2. CHAPTER 7: Results

20.1.1. SECTION 7.1: Quantitative Results

Table 20.1 and fig. 21.1 (overleaf) report the results of various traceback analyses, whose parameters are described below.

Method_A is the one adopted for the analysis and its details have been fleshed out throughout chapter 6. It considers a putative precedent as an available unit when it is attested at least twice in the main corpus. In order for a schema to be created, the putative slot must be instantiated by at least two different fillers. A unit that is attested in the main corpus with at least two different internal word orders (WO) is considered to be a packet. A unit that is attested in the main corpus with only one internal WO is considered to be a unit whose word order is fixed.

Method_B is identical to method A, with the exception that, in order for a putative precedent to be considered as an available unit, it must be attested at least three times in the main corpus.

Method_C considers a putative precedent as an available unit if it is uttered by the child (and the child only) at least once (excluding imitations and self-repeats) in the main corpus. Since a unit has to appear only once, a slot is created whenever the target sentence and its precedent share some kind of lexically-specific material, but differ in one or two elements which share the same morphological and/or semantic proprieties (as in Lieven et al. (2009), discussed in 3.5 and 6.4). Hence, the slot need not be instantiated by two different fillers. WO variability is assumed and need not be found in the main corpus.

Method_D is identical to Method_C, but the main corpus instantiations of a component unit can be uttered by any speaker (not only the child).

Table 20.1: Results: using different traceback methods to analyse Roberto's target sentences. Method_A is the method adopted for the main analysis. For each method, the frequency threshold a precedent had to meet in order to be considered as an available unit and who could have uttered it are reported in the "method's description". Hence, in Method_C, target sentences were traced back to strings that the child (and the child only) uttered in the main corpus at least once (excluding imitations and self-repeats).

METHOD	method's description	no. of 0-operation derivations	no. of 1-operation derivations	no. of 2-operation derivations	no. of 3-operation derivations	no. 4-or-more-operation derivations
Method_A	2 any speaker	113	219	142	90	69
Method_B	3 any speaker	94	205	135	101	71
Method_C	1 child only	109	216	131	91	55
Method_D	1 any speaker	144	252	140	99	59
METHOD	**method's description**	**no. of lexical fails**	**no. of constructional fails**	**no. of both constructional and lexical fails**	**total fails**	**total sentences**
Method_A	2 any speaker	38	90	7	135	768
Method_B	3 any speaker	48	106	8	162	768
Method_C	1 child only	68	89	9	166	768
Method_D	1 any speaker	24	47	3	74	768

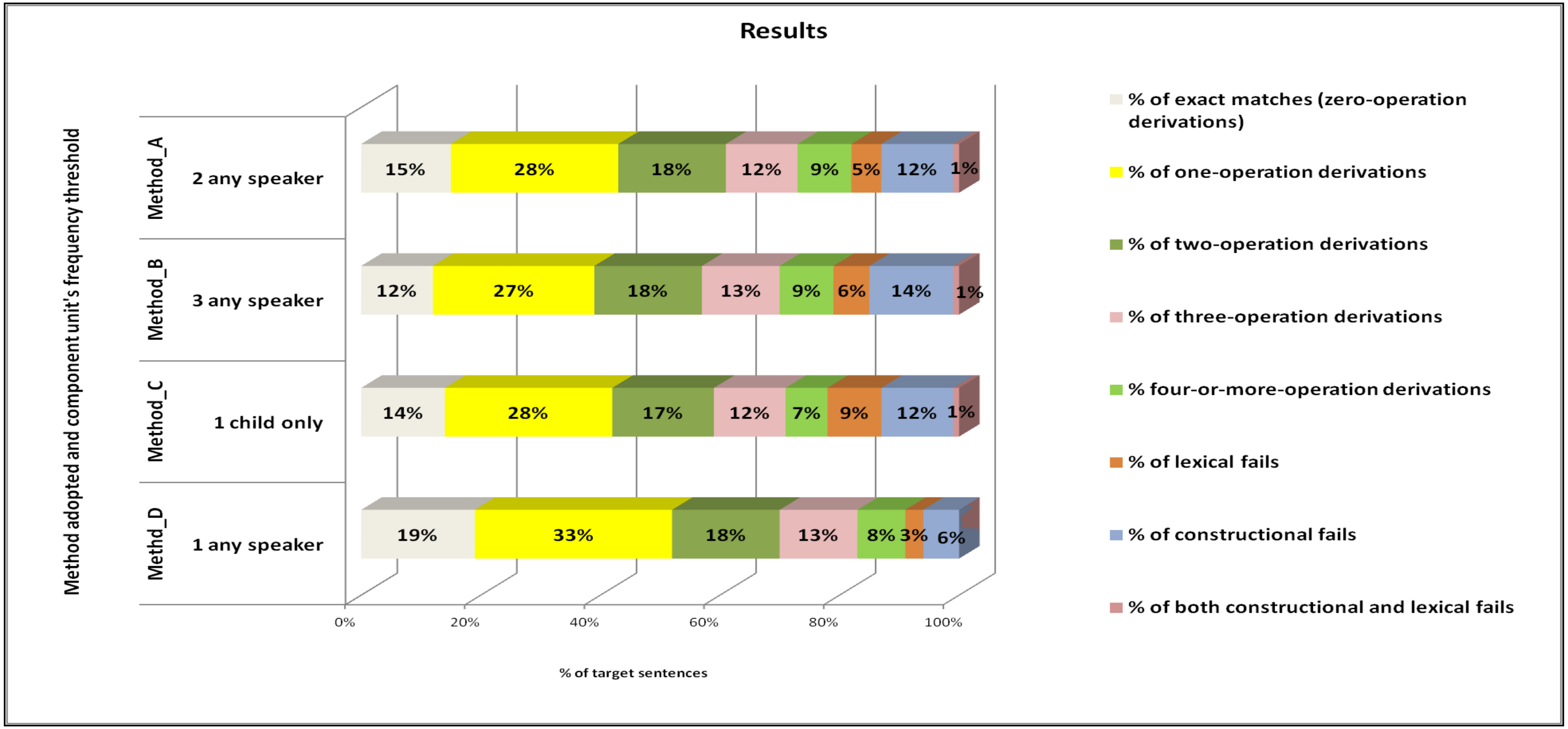


Figure 20.1: Comparing the results of different traceback methods. Method_A is the method adopted for the main analysis. For each method, the frequency threshold a precedent had to meet in order to be considered as an available unit and who could have uttered it are reported on the left hand-side of each bar. Hence, in Method_C, target sentences were traced back to strings that the child (and the child only) uttered in the main corpus at least once (excluding imitations and self-repeats).

20.2.2. SECTION 7.2: The Longest Derivation

Target sentence SB180 (1) required 10 superimpositions to be derived from its putative component units (reported in fig. 20.2) and it represents the derivation (fig. 20.3-20.8) that required the largest number of operations.

1. **CHI: <sai> [/] sai che io (.) ho fatto una corsa (..) grande + e sono arrivato (.) dal Luca +.+ e ho fatto puff.*
   sai che io ho fatto una corsa grande e sono arrivato dal Luca e ho fatto puf.

sa-i	*che*	*io*	*ho*	*fatto*	*un-a*
know-PRS.2.SG	that	I	have(PRS.1.SG)	done	a-F.SG

cors-a	*grand-e*	*e*
run-F.SG	big-SG	and

sono	*arriv-a-t-o*	*da=l*
be(PRS.1.SG)	arrive-TV(conj.I)-PTCP-M.SG	at=the(M.SG)

Luca	*e*	*ho*	*fatto*	*puff.*
Luca	and	have(PRS.1.SG)	done	puff.

   ("You know that I have done a big run and I have arrived at the Luca and I've done ʽpuff'?").
   "Did you know that I ran a long way and got to Luca and went ʽpuff' ?".
   (week5.2014.02.18.B: lines 1423 and 1425)

Since target sentence SB180 is made of one main clause (*did you know that*, lit. *you know that*) and three co-ordinated subordinate clauses, its derivation is presented by firstly showing the derivations of the three subordinate clauses separately (***I ran a long way*** (lit. *I've done a big run*; fig. 20.3) + ***I got to Luca*** (lit. *(I) am arrived at the Luca*; fig, 20.4) + ***I went "puff"*** *(*lit. *(I) have done "puff"*; fig. 20.5) and how they are conjoined together (fig. 20.6 and 20.7). Successively, the resulting clause (*I ran a long way and got to Luca and went "puff"*) is used to fill the slot *SENTENCE* of the main clause (fig. 20.8) yielding the final target sentence. Note that it is plausible that the utterance has been put together following sequences other than the ones reported; the method makes no assumptions as to the order in which component units are assembled. In each figure, superimpositions are highlighted in yellow and orange, while lexical material shared by the units superimposed is highlighted in blue.

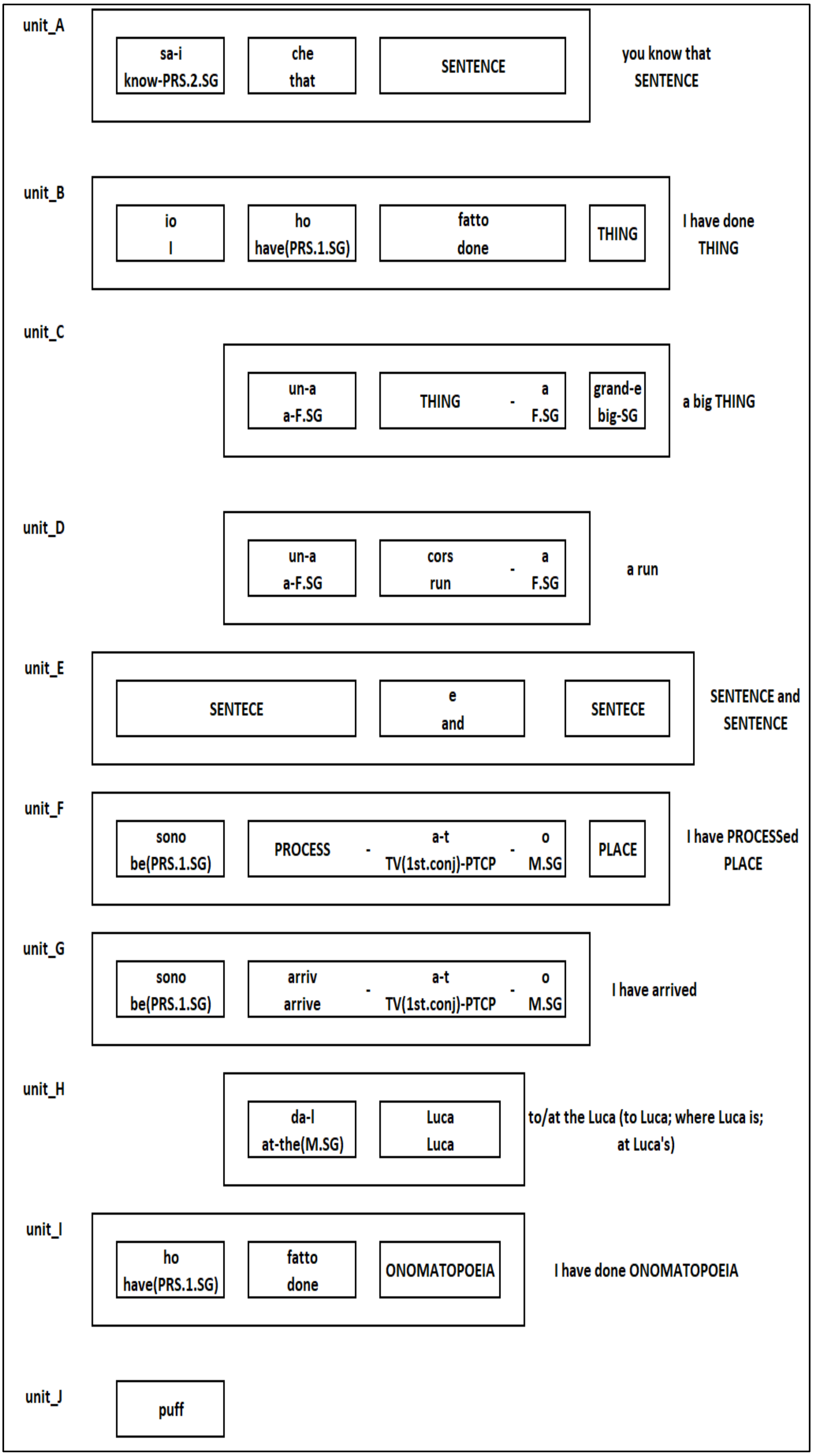


Figure 20.2: The putative precedents of target sentence SB180 (1).

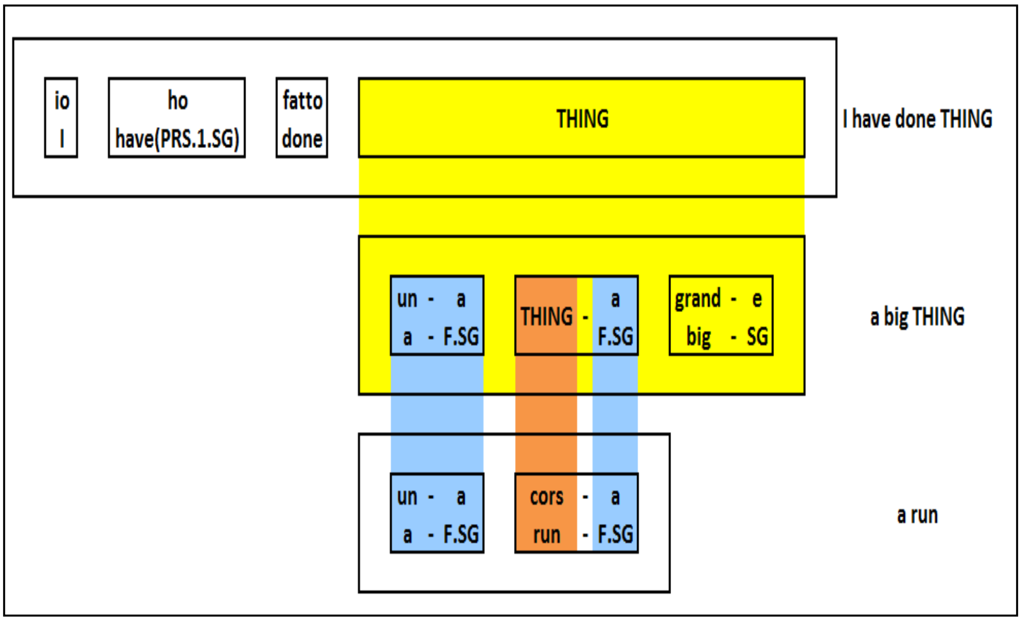


Figure 20.3: Deriving target sentence SB108 (1); 1 of 6. ***Io ho fatto una corsa grande*** **"I ran a long way" (lit.** ***I have done a big run*****).**

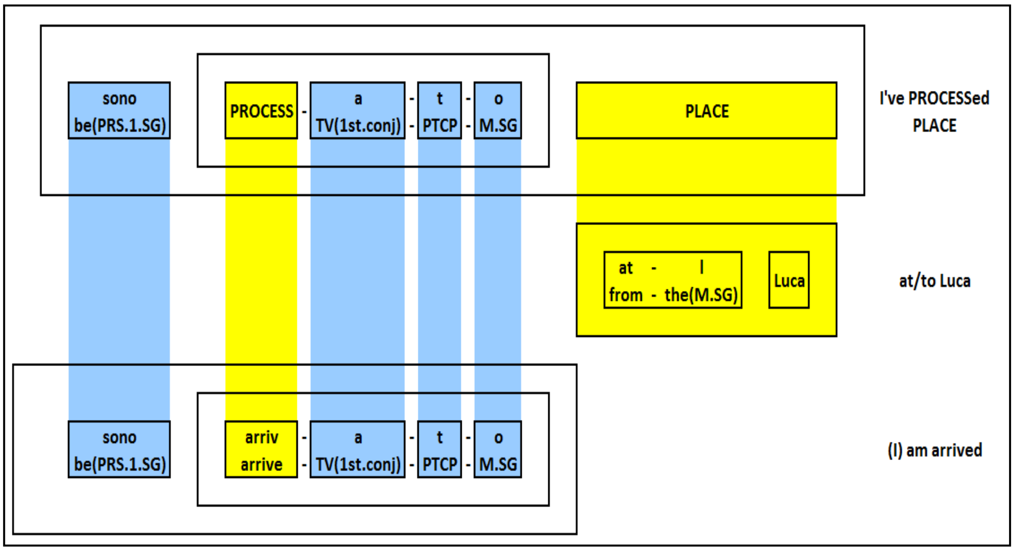


Figure 20.4: Deriving target sentence SB108 (1); 2 of 6. ***Sono arrivto dal Luca*** **"I got to Luca" (lit.** ***(I) am arrived at the Luca*****).**

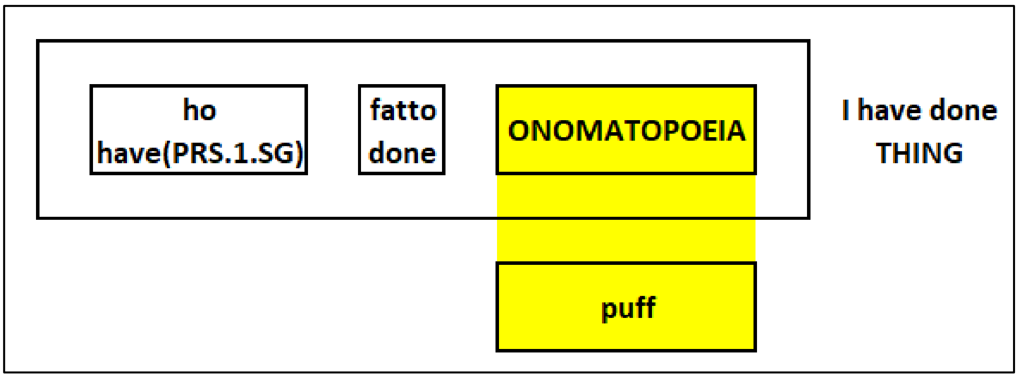


Figure 20.5: Deriving target sentence SB108 (1); 3 of 6. ***Ho fatto "puff"*** **"I went 'puff' " (lit.** ***(I) have done "puf"*****).**

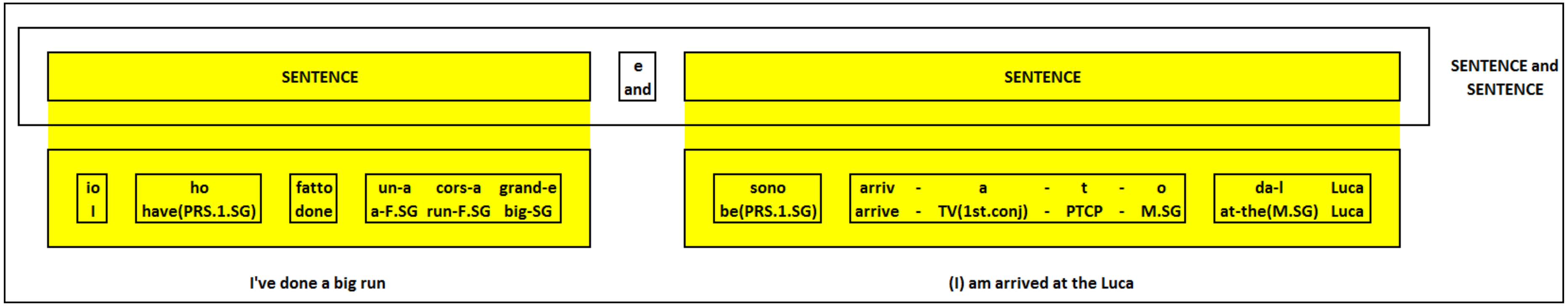


Figure 20.6: Deriving target sentence SB108 (1); 4 of 6. *Io ho fatto una corsa grande e sono arrivato dal Luca* "I ran a long way and got to Luca" (lit. *I have done a big run and am arrived at the Luca*).

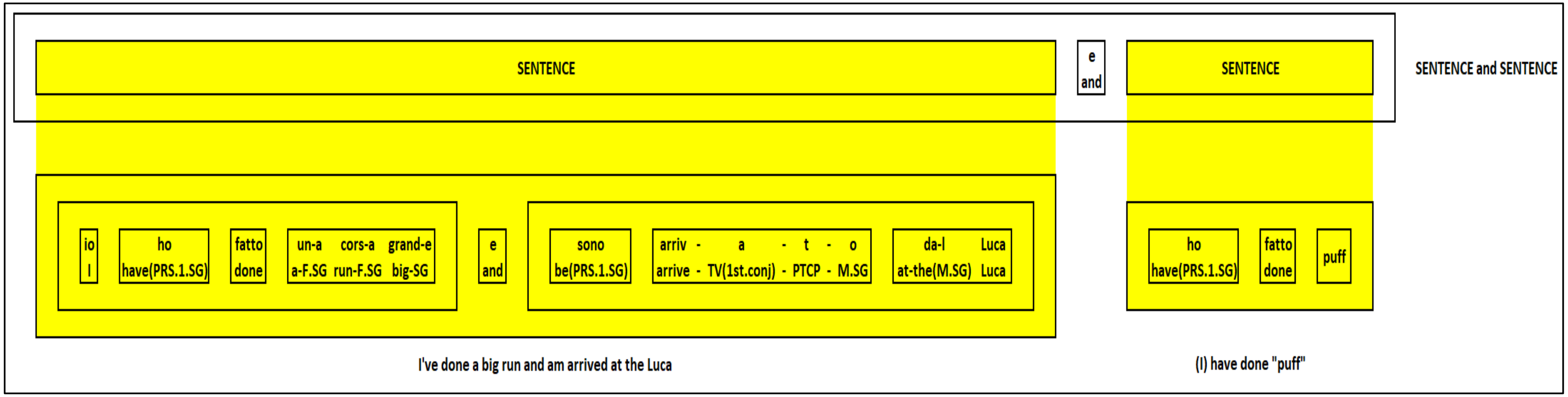


Figure 20.7: Deriving target sentence SB108 (1); 5 of 6. *Io ho fatto una corsa grande e sono arrivato dal Luca e ho fatto "puff"* "I ran a long way and got to Luca and went 'puff' " (lit. *I have done a big run and am arrived at the Luca and have done "puff")*.

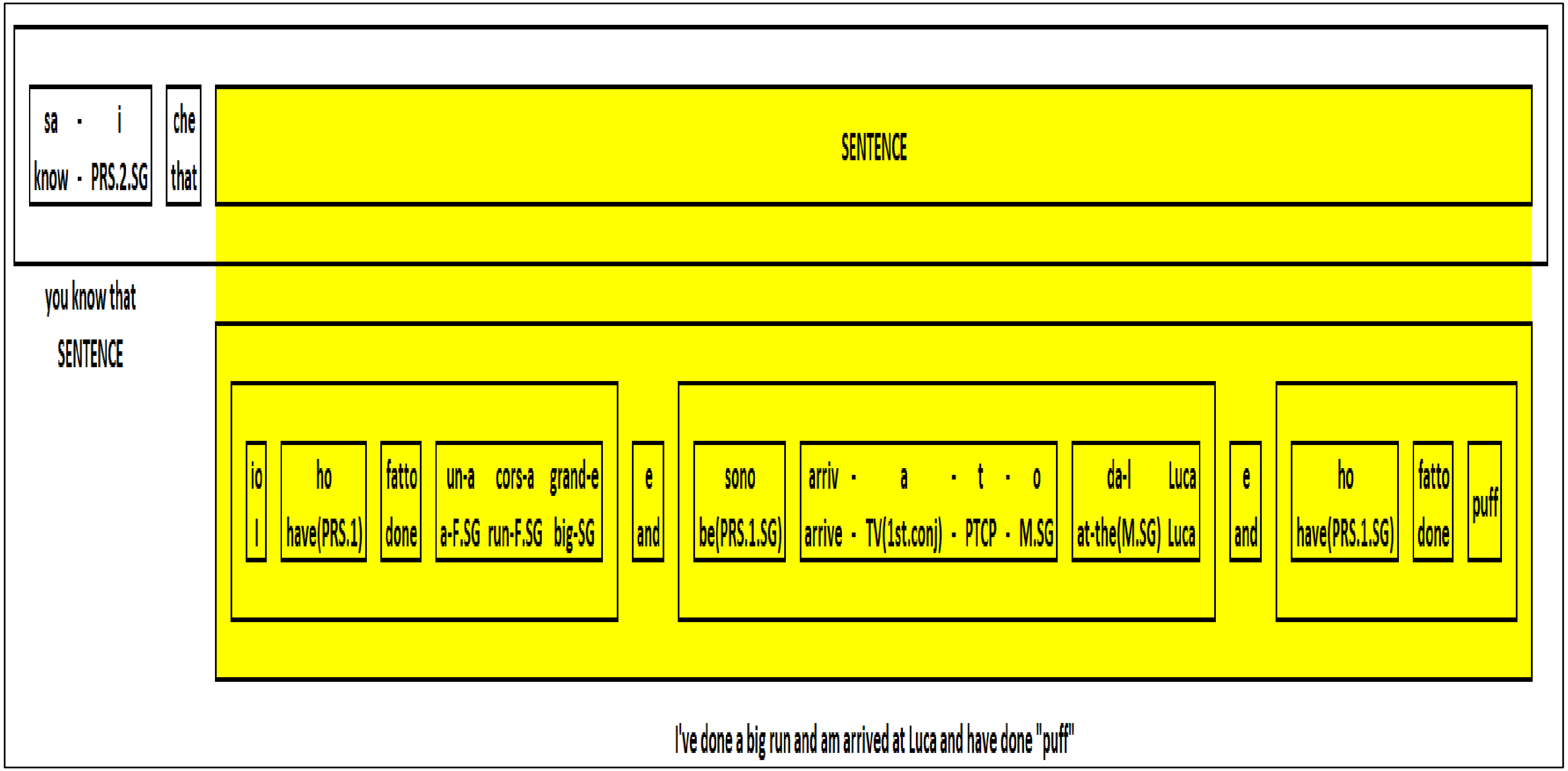


Figure 20.8: Deriving target sentence SB108 (1); 6 of 6. ***Sai che io ho fatto una corsa grande e sono arrivato dal Luca e ho fatto" puff"*** **"Did you know that I ran a long way and got to Luca and went 'puff' " (lit.** ***you know that I have done a big run and am arrived at Luca and have done "puff"*****).**

20.2.3. SECTION 7.2: Qualitative Results (Fully Lexically-Specific Strings)

The method identified 502 Fully Lexically-specific Strings (fig. 20.9, table 20.2), 186 were fixed strings which could not possibly present other orders (such as article-noun combinations) and only 6 (1%) were packets.

Table 20.2: Types of Fully Lexically-specific Strings identified.

Types of Fully Lexically-specific Strings	no.
fixed-strings whose internal order is the only possible grammatical order	186
single words	156
fixed-strings	147
onomatopoeias	7
packets	6
total	**502**

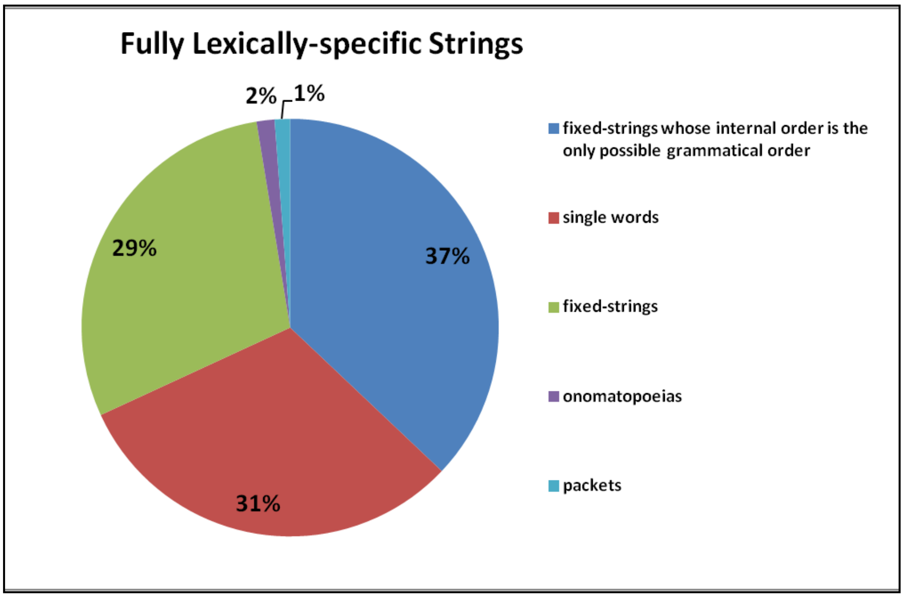


Figure 20.9: types of Fully Lexically-specific Strings.

Fully Lexically-specific Strings could be of various lengths in terms of the numbers of words and morphemes. Tab. 20.3 presents the longest Fully Lexically-specific Strings (FLSS in tables 20.3-205) in terms of number of morphemes, subsequently ordered by number of words and frequency of retrieval[5]. Tab. 20.4 orders the same type of units by number of words, number of morphemes and frequency of retrievals. Finally, tab. 20.5 gives the most retrieved strings, subsequently ordered by number of morphemes and number of words.

Table 20.3: the longest Fully Lexically-specific Strings (FLSS), ordered by number of morphemes, number of words and frequency of retrieval.

FLSS	translation	no of morphemes	no of words	frequency of retrieval
c'era una volta una bella favolina	once upon a time there was a little fairy tale	14	7	1
questo è il registratore del Luca	this is Luca's voice recorder	9	6	2
dove è andato l'altro	where has the other one gone?	8	5	1
a casa della nonna	at grandma's home	8	3	2
una bella favolina	a nice little fairy tale	8	3	1

Table 20.4: the longest Fully Lexically-specific Strings (FLSS), ordered by number of words, number of morphemes and frequency of retrieval.

FLSS	translation	no of words	no of morphemes	frequency of retrieval
c'era una volta una bella favolina	once upon a time there was a little fairy tale	7	14	1
questo è il registratore del Luca	this is Luca's voice recorder	6	9	2
dove è andato l'altro	where has the other one gone?	5	8	1
cosa c' è dentro qua?	what's inside here?	5	5	1
cosa c'è dentro qui	what's inside here?	5	5	1

Table 20.5: the most retrieved Fully Lexically-specific Strings (FLSS), ordered by frequency of retrieval, number of morphemes and number of words.

FLSS	translation	no. of retrievals	no. of morphemes	no. of words
questo	(quest-o/this-M.SG)	13	2	1
il registratore	(il registrator-e/the voice.recorder-M.SG)	10	3	2
quello	(quell-o/that.one-M.SG)	8	2	1
pastasciutta	(pastasciutt-a/pasta-F.SG)	6	2	1
la macchina	(l-a macchin-a/the-F.SG car-F.SG)	5	4	2
la moto	(l-a mot-o/the-F.SG motorbike(F.)-SG)	5	4	2
Luca	(Luca)	5	1	1

The most frequent Fully Lexically-specific String is *quest-o* (*this-M.SG*). The longest one is reported in the main text, Volume I, section 7.2.1, sentence (50).

20.2.4. SECTION 7.2: Qualitative Results (Schemas with Slots)

The method identified 698 schemas with slots, 506 (74%) containing only one slot and 182 (26%) being two-slot schemas. Such schemas could be at the word (single word schemas) or at the clause (multi-word schemas) level. In the latter case, they could be either schematic-packets or fixed-schemas (fig. 20.10).

[5] i.e., the number of times Roberto uttered those strings in the **test corpus**.

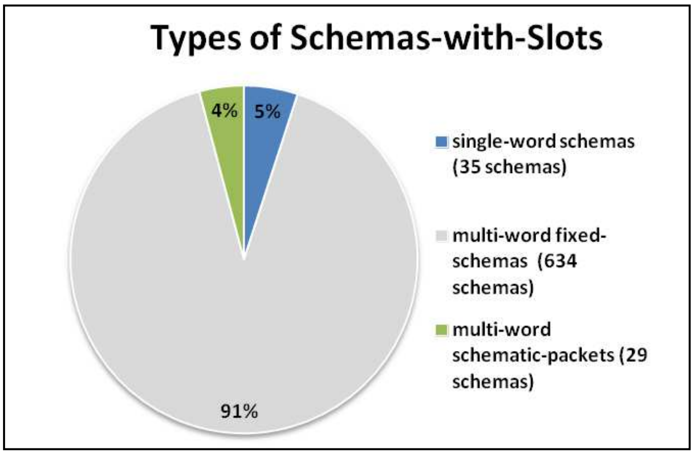


Figure 20.10: types of schema.

20.2.4.1. Semantic Slots

There are 10 types of semantic slots: *THING*, *PROCESS*, *PLACE*, *SENTENCE*, *QUALITY*, *PERFORMANCE*, *CONSEQUENCE, PROPERTY, DIRECTION* and *QUESTION*. Table 20.6 and fig. 20.11 show that the most attested semantic slots are *THING*, *PROCESSS*, *PLACE* and *SENTENCE*, which together account for 96% of all semantic slots. A few examples of these very frequent semantic slots are given below. For the sake of brevity, only one or two instantiations of the slot (and the schema) illustrated are provided.

Table 20.6: Types of semantic slots.

SLOT TYPE	no.
THING	426
PROCESS	230
PLACE	52
SENTENCE	51
QUALITY	14
PERFORMANCE	10
CONSEQUENCE	3
PROPERTY	2
DIRECTION	2
QUESTION	1
total	791

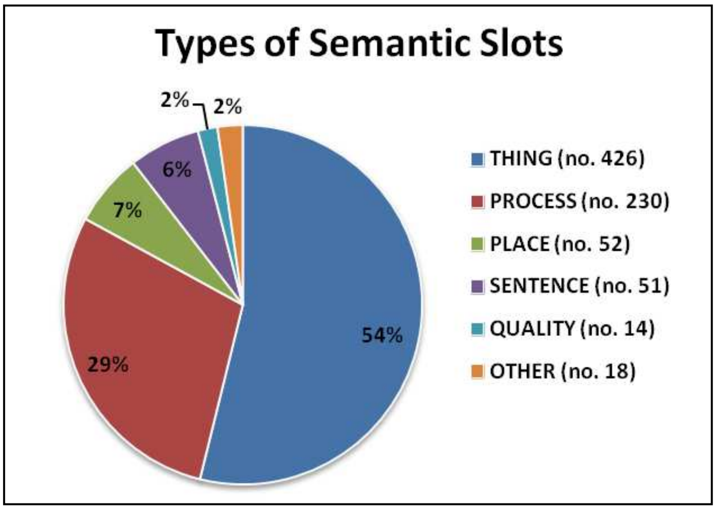


Figure 20.11: types of semantic slots.

Examples of ***THING*** **slots** can be found in the main test (Volume I, section 7.2.2.1)

PROCESS **slots** are generalisations across verbs and full VPs. There are four subtypes of *PROCESS* slots:

a) PROCESS

   These are generalisations across various verbs that can take various morphological endings and can also be generalisations across whole VPs and constructions (*transitives*, *intransitives* and so on; fig. 20.12a)

b) IRREGULAR

   These are generalisations across various instances of a specific irregular verb. Since irregular verbs often present irregular and multiple roots (e.g. suppletion), their roots cannot be phonologically specified and hence they represent semantic generalisations. However, their scope is narrower than *PROCESS* slots, as the latter can be filled by any verb and VP, whereas *IRREGULAR* slots can be filled only by occurrences of a specific verb. They are reported as *AVERE* "to have", *ESSERE* "to be", *ANDARE* "to go" and so on (fig. 20.12b)

c) SEMANTICALLY_NARROW_PROCESS

   These are generalisations across various occurrences of a specific verb (such as *succedere* "to happen") and can be instantiated by only that specific verb and the various forms it can take (both composite and simple forms; fig. 20.12c)

d) PROCESS-specific.inflection

These are generalisations across verbs' roots whose morphological endings (small letters) are specified. Hence, they are part of schemas that specify aspect, mood, tense, person, number and gender of their verbs (fig. 20.12d), as well as any clitic pronoun that might merge with them.

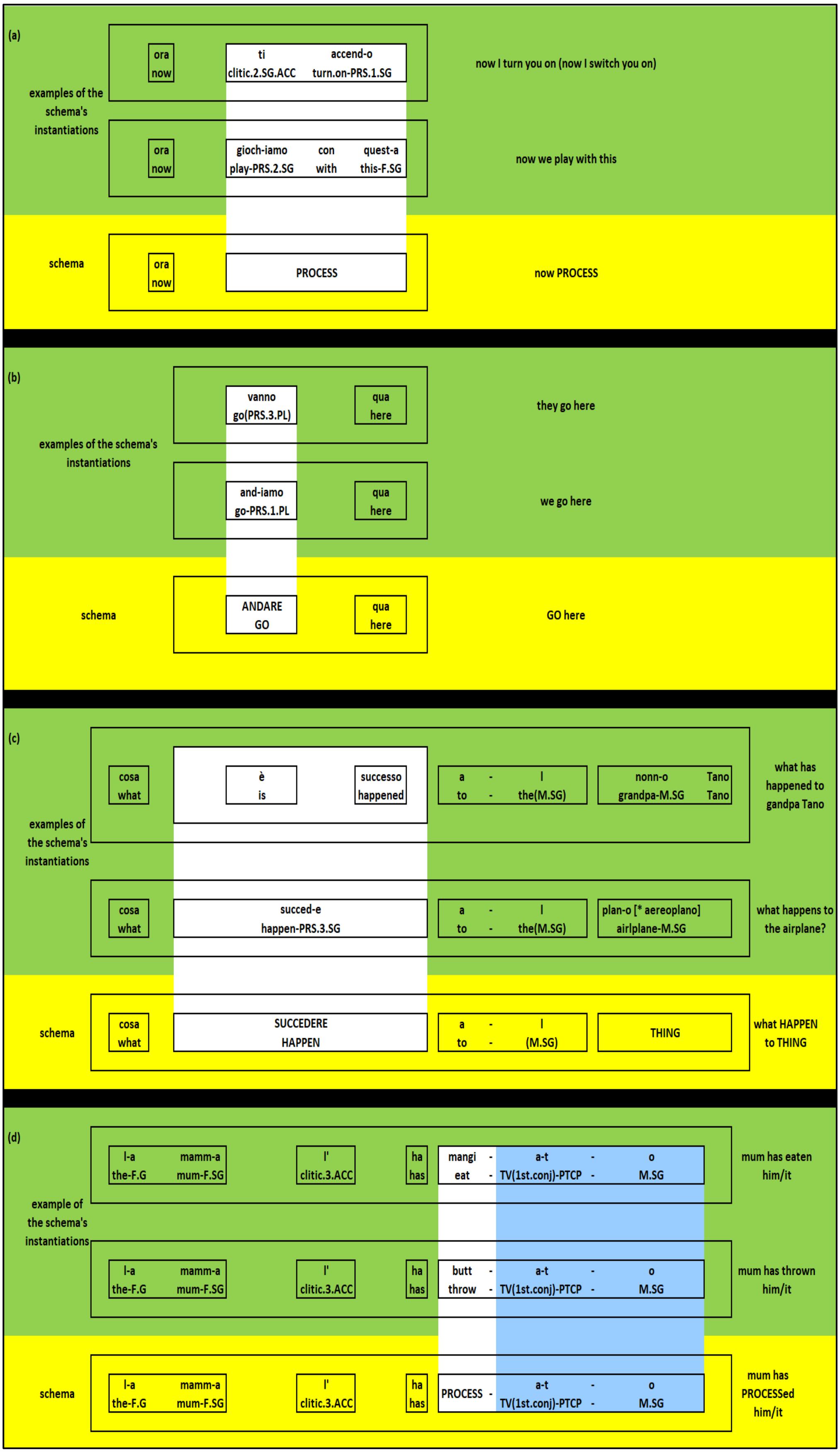


Figure 20.12: the PROCESS slot (highlighted in white). Schemas are highlighted in yellow, whereas their instantiations are highlighted in green. Relationships of elaboration (slot formation) are highlighted in white and shared lexical material is highlighted in blue.

***PLACE* slots** are generalisations across lexical items that represent location in a particular schema. They are overwhelmingly instantiated by PPs (fig. 20.13a).

***SENTENCE* slots** are generalisations across whole sentences. Such kinds of slots are normally bound to vocatives (*mum SENTENCE*), conjunctions (*SENTENCE and SENTENCE*) and certain fixed phrases Roberto constantly uses to draw adults' attention (fig. 20.13b)

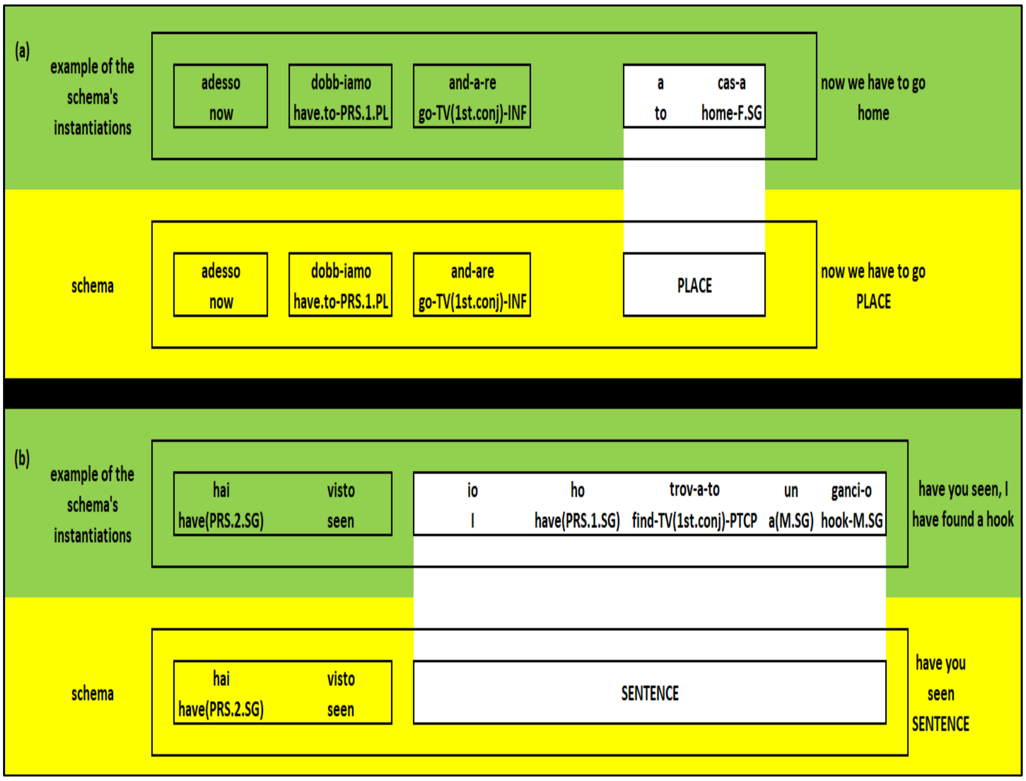


Figure 20.13: PLACE (a) and SENTENCE (b) slots (in white). Schemas are highlighted in yellow and their instantiations are highlighted in blue.

20.2.4.2. Morphological Slots

The method yielded four types of morphological slots: *NOUN.INFLECTION*, *GENDER.NUMBER*, *VERB.INFLECTION* and *THEMATIC VOWEL* (table 20.7, fig. 20.44).

Table 20.7: Types of morphological slots.

SLOT TYPE	no.
process(root)-INFLECTION	54
GENDER.NUMBER	17
noun-INFLECTION	6
THEMATIC VOWEL	2
total	79

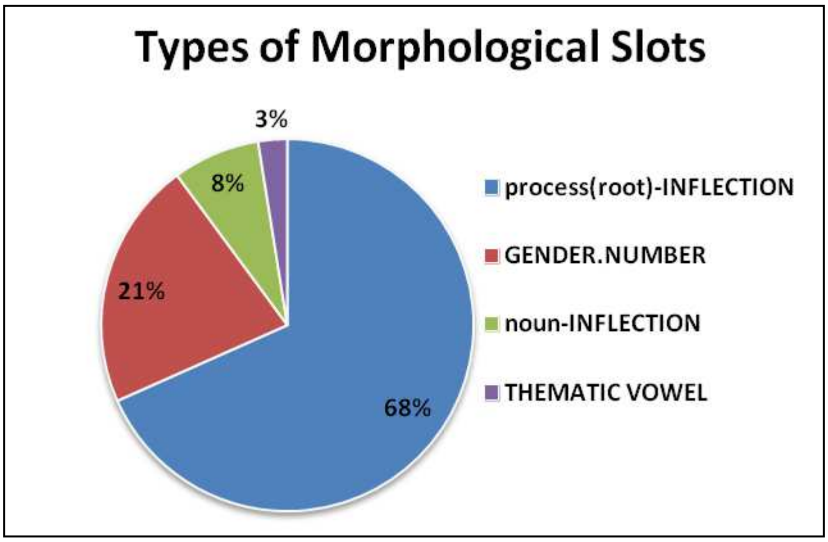


Figure 20.14: types of morphological slots.

For examples of process(root)-INFLECTION slots, refer to the main test (Volume I, section 7.2.2.2)

GENDER.NUMBER slots are generalisations across gender and number markers (*-o*, *-i*, *-a*,*-e*) that are created when the schemas to which they belong allow flexibility on gender and number of an otherwise fixed element. These slots can be bound to specific instances of adjectives, past participles (fig. 20.15a), clitic pronouns (fig. 20.15b) and so on.

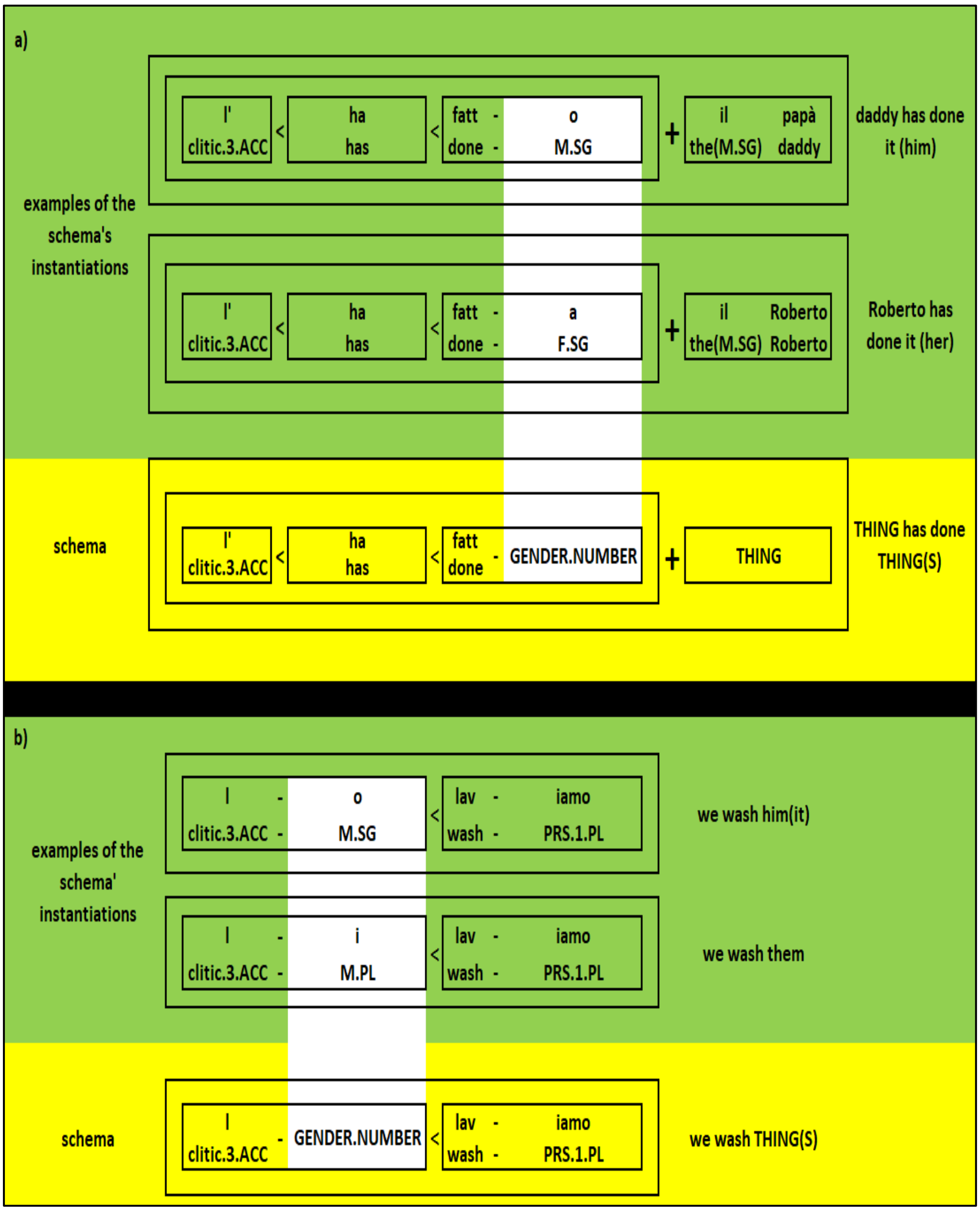


Figure 20.15: GENDER.NUMBER slot (in white); "<" indicates (fixed) linear order, whereas "+" indicates that the sequence is flexible. Hence, the schema (yellow strip) in (a) indicates that the sequence "*l'ha fatt-GENDER.NUMBER*" can either follow or precede the slot *THING*.

20.2.4.3. Most retrieved Schemas

Table 20.8 reports the 10 most retrieved schemas: all but one present only one slot; they are also all schemas whose slot is *SENTENCE* and whose internal word order is fixed. Overall, these very frequent schemas account for 226 operations (15% of the 1087 operations yielded by the method). The most frequent schema which does not present a *SENTENCE* slot is *quello è un THING* "that is a THING" (fig. 20.16) and it was retrieved 6 times.

Table 20.8: the most retrieved schemas.

schema	translation into English	no. of retrievals
Luca SENTENCE	Luca SENTENCE	27
SENTENCE Luca	SENTENCE Luca	22
certo SENTENCE	of course SENTENCE	17
no SENTENCE	no SENTENCE	17
dai Luca SENTENCE	c'mon Luca SENTENCE	16
e SENTENCE	and SENTENCE	15
guarda SENTENCE	look SENTENCE	15
SENTENCE e SENTENCE	SENTENCE and SENTENCE	15
allora SENTENCE	hence SENTENCE	11
dai SENTENCE	com'on SENTENCE	11

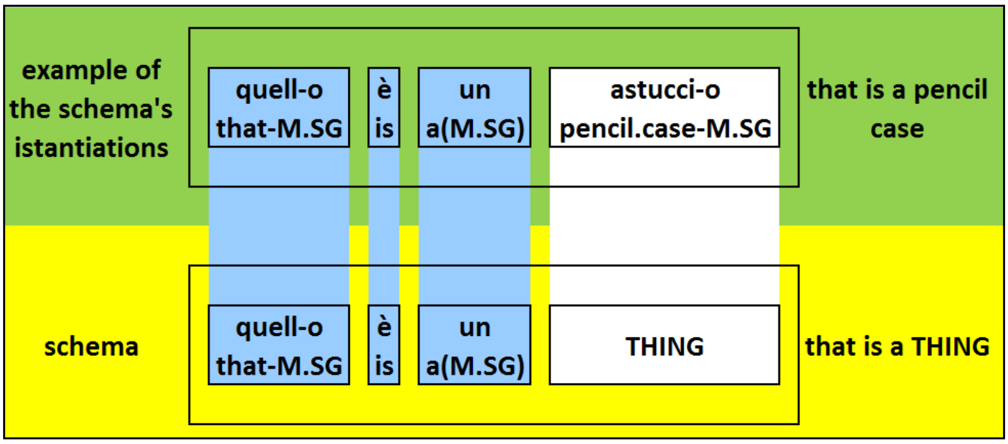


Figure 20.16: the schema *that is a THING* (yellow strip) and one of its instantiations (green strip). Fixed lexical material is in blue, the slot is highlighted in white.

20.3. CHAPTER 8: Analysis

20.3.1. SECTION 8.1: *Mettere* "to put" and the Caused-Motion-Construction

Fig 20.17 reports four schemas built around *mettere* "to put", which instantiate the Caused-Motion-Construction (cmc). Note that those schemas do **not** represent all the instantiations of the cmc built around *mettere* "to put" attested in the dataset. Rather, fig. 20.7 reports schemas that are instantiated twice or more in the main corpus **and** that were used to successfully analyse Roberto's target sentences.

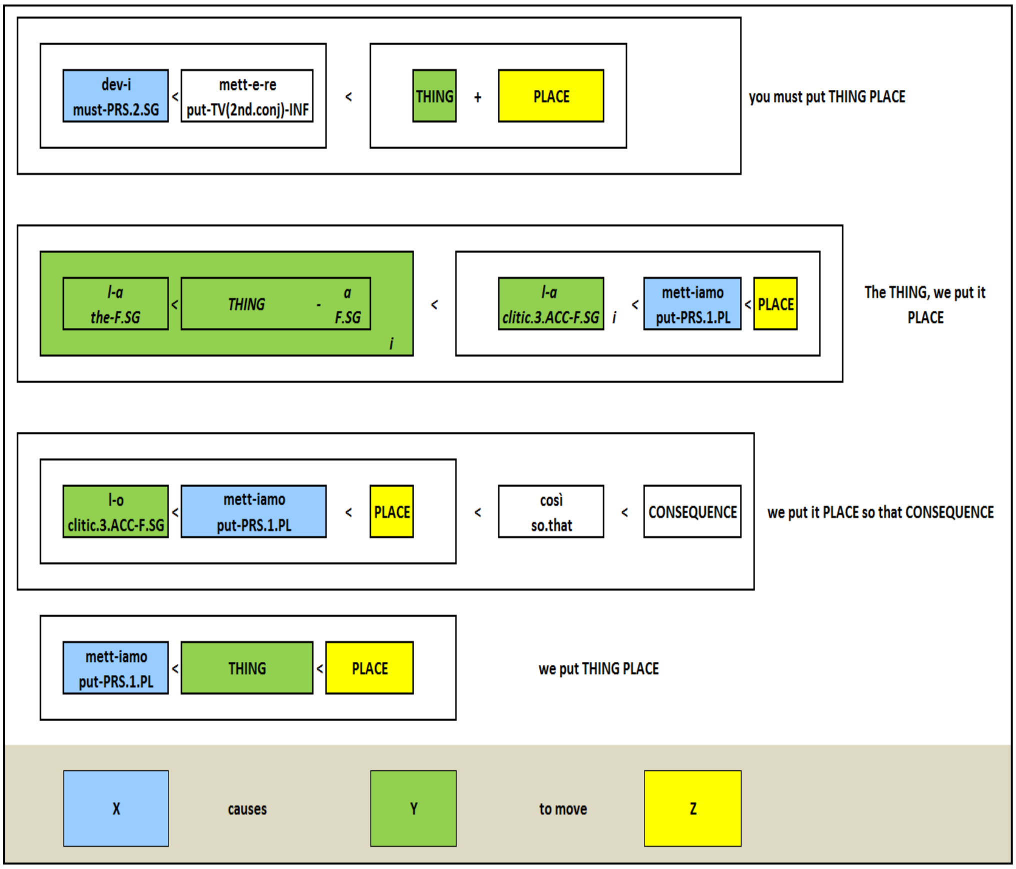


Figure 20.17: schemas instantiating the caused-motion construction built around *mettere* "to put" which were retrieved in the test corpus. The symbol "<" indicates linear sequence, whereas "+" indicates that the elements have no fixed word order (refer back to the caption of fig. 20.15). Elements in *Italics* are co-indexed.

20.3.2. SECTION 8.4.1: Phonological Mistakes, omissions and ill-imitations

Four sentences are classified as Hard_Constructional_Fails because the child seems to mispronounce a word, these mispronunciations happen to be existing words. The result is a target sentence that cannot be traced back. In target sentences SB315 (2), the likely target word *ma* (/ma/ "but") is pronounced as *da* (/da/ "from"). The putative child's target sentence (3) could be traced back. However, the actual target utterance (2) cannot be accounted for.

2. *CHI:* ***da** quello è un-o spazzaneve*
da *quell-o* *è* *un* *spazzanev-e*
from that-M.SG is a(M.SG) snowplough(M.)-SG
“***from** that is a snowplough.”
(week6.2014.02.18.B: line 2367)

3. ***Ma** quello è un-o spazzaneve*
ma *quell-o* *è* *un* *spazzanev-e*
but that-M.SG is a(M.SG) snowplough(M.)-SG
“**but** that is a snowplough.”

Four fails are caused by Roberto dropping an element from an attested schema. Fig. 20.18 shows target sentence SA109 (fig. 20.18a) and the putative child’s target (20.18b). The two differ in that Roberto drops the preposition *a* “to” on which the subordinate depends. Fig. 20.18b could have been accounted by the schema in fig. 20.18c. Roberto’s sentence (20.18a) cannot.

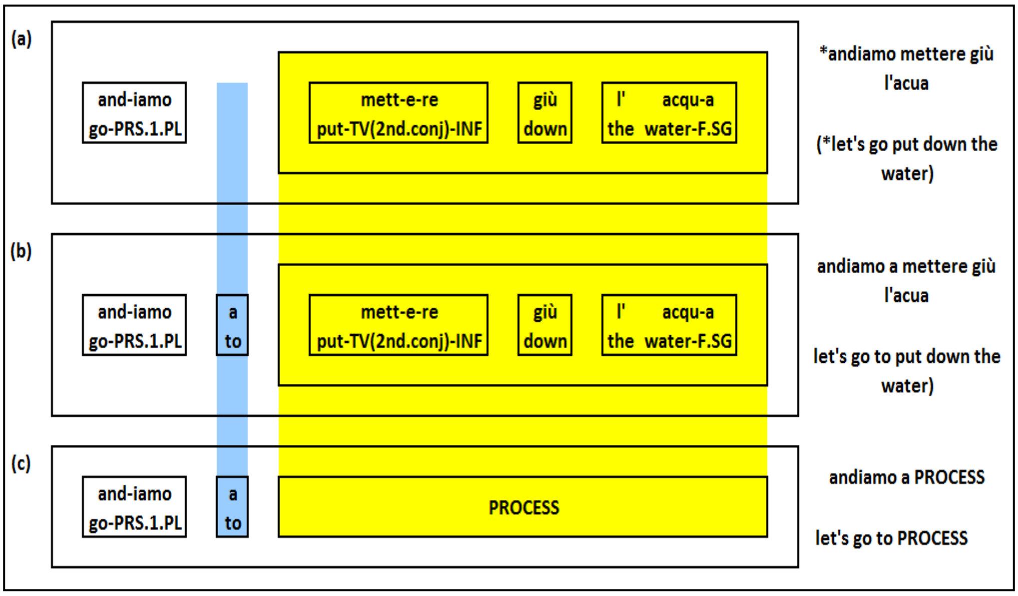


Figure 20.18: Target sentence (SA109) (a), the putative sentence the child tried to utter (b) and the schema that could explain it (c). The element dropped is highlighted in blue. Relationships of elaboration are in yellow.

One sentence is an ill-imitation and is reported in the main text (8.4.1).

20.3.3. SECTION 8.5: t-units analysis

20.3.3.1. Identifying T-units

Firstly, t-units were identified within target sentences. A t-unit was defined in both semantic and grammatical terms. Semantically, it had a meaning that was complete and independent, that is, t-units had to be semantically coherent so that their meaning could be (mostly) inferred without the support of other sentences. Grammatically, a t-unit had a main clause (with its attached and/or embedded subordinate clauses and non-clausal structures) that presented at least one verb inflected for person (1^{st}, 2^{nd}, 3^{rd}) and number (plural vs. singular). Hence, target sentence SB306 (1) consists of a single t-unit. Similarly, target sentence SB180 (2) is a single t-unit made of a main clause (*did you know that*) and three subordinate classes (*I ran a long way* - *and got to Luca* - *and went "puff"*). In the examples in the following pages, t-units are delineated by slashes (/).

1. /N*o però sei troppo piccola per entrare nella mia casetta/*
/no, però sei troppo piccol-a per
/no , but be(PRS.2.SG) too small-F.SG to
entr-a-re ne=ll-a mi-a
enter-TV(conj.I)-INF in=the-F.SG my-F.SG
cas-ett-a/
house-little(ENDER)-F.SG/
"No, but you are small big to enter into my little house."
(week5.2014.2.18.B: lines 2334-2335)

2. *<sai> [/] sai che io (.) ho fatto una corsa (..) grande + e sono arrivato (.) dal Luca +.+ e ho fatto puff.*
/sai che io ho fatto una corsa grande e sono arrivato dal Luca e ho fatto puf./
/sa-i che
know-PRS.2.SG that
io ho fatto un-a cors-a grand-e
I have(PRS.1.SG) done a-F.SG run-F.SG big-SG
e sono arriva-t-o
and be(PRS.1.SG) arrive-TV(conj.I)-PTCP-M.SG
da=l Luca
at=the(M.SG) Luca
e ho fatto puff /.
and have(PRS.1.SG) done puff.
("You know that I have done a big run and I have arrived at the Luca and I've done ʽpuff'?").
"Did you know that I ran a long way and got to Luca and went ʽpuff'?".
(week5.2014.02.18.B: lines 1423 and 1425)

Overall, 113 sentences that either did not present a verb inflected for person and number or were verbless were excluded from the analysis as they did not contain any t-unit. A further 31 sentences were classified as lexical fails and were therefore excluded as the purpose of the analysis was to compare successful derivations and constructional fails.

20.3.3.2. Co-ordinated structures

Target sentences composed of co-ordinated clauses could be analysed as comprising either several t-units or one single t-unit. Hunt (1965) and Nippold et al. (2005) used the presence/absence of both a [+FINITE] verb and a subject as a benchmark for identifying t-units. According to their method, a sentence like (3) contains two t-units, as both clauses contain a subject and a [+FINITE] verb. However, (4) contains only one t-unit as the two clauses share a co-ordinated subject.

3. /Claire bought a bike/ and she cleaned her house/.
4. /Claire bought a bike and cleaned her house /.

However, such a method of delineating t-units made of co-ordinated clauses (4) and co-ordination between different t-units (3) did not seem appropriate when dealing with the pro-drop nature of Italian. Hence, when the target sentence contained two (or more) co-ordinated clauses presenting a verb inflected for person and number which shared the same overt subject, this was considered a single t-unit (5). When both clauses were subjectless (6) or had different subjects (either overt or null) the sentence was analysed as containing two (or more) different t-units.

5. *E tu tieni questa e scrivi*

/e	*tu*	*tien-i*	*quest-a*
and	you(2.SG.NOM)	take-PRS.2.SG	this-F.SG
e	*scriv-i/*		
and	write-PRS.2.SG		

"And you take this one and write."
(week5.2014.2.18.B: lines 2492)

6. *È bravissimo e può giocare*

/è	*brav-issim-o*	/	*e*	*può*
is	good-SUPERLATIVE-M.SG	/	and	can(PRS.3.SG)
gioc-a-re				
play-TV-(conj.I)-INF				

"(He) is very good and can/may play."
(week5.2014.2.18.B: lines 2150)

20.3.3.3. Subordination and Juxtaposition

Another issue that arose while identifying t-units was Roberto's reliance on attention drawing strategies which generally took the form of imperatives. For instance, many target sentences start with the formula *guard-a SENTENCE* "look-IMP.2.SG SENTENCE".

A choice was made that those target sentences would be considered a single t-unit whenever a complementiser was attested (7). When the two clauses were juxtaposed (8) the target sentence was considered as presenting two (or more) t-units.

7. *<Luca> [<] +. guarda cosa ho fatto.*
   /Luca, guard-a cosa ho fatto/
   Luca, look-IMP.2.SG what have(PRS.1.SG) done
   "Luca, look at what I've done."
   (week6.2014.02.18.B: lines 1227 and 1229)

8. *<guarda, eh@i, > [<] ti ho portato questo.*
   /guard-a, / / ti ho
   look-IMP.2.SG, / clitic.2.SG.DAT have(PRS.1.SG)
   port-a-to quest-o /
   bring-TV(conj.I)-PTCP this-M.SG /
   "Look, I brought you this one."
   (week6.2014.02.18.B: line 582)

Such a procedure yielded 46 one-or-two-word t-units that instantiated four t-unit types: *mamma + guard-a* "mamma + look-IMP.2.SG", *guard-a* "LOOK-IMP.2.sg", *hai visto* "have(PRS.2.SG) seen" and *sent-i* "hear-IMP.2.SG". Because such strings/t-units were derivable and comprised only a single word, they had the effect of drastically skewing the mean length of successful derivations. The solution was to factor in only one instance for each t-unit. That is, once target sentences had been broken down into t-units, the analysis considered only the t-unit **types** produced by Roberto.

Once verbless sentences, lexical fails and redundant instances were discarded, there was a total of 646 t-units: 90 were contained in constructional fails and 556 were contained in successful derivations.

For the coding of what was considered a clause within a t-unit, refer to the main test (section 8.5).

21.

APPENDIX_III:

An Experimental Insight into the Development of Morpho-Syntactic Competence

Page intentionally left blank

21.1. THE ADULT PARTICIPANTS

Table 21.1: The adults who contributed answers for the experimental dataset.

gender	age	other languages spoken (in order of fluency)	profession	level of instruction
m	31	Spanish, English	student, tourist guide	undergraduate degree
f	25	English	farmacist	postgraduate degree
f	25	English	student	undergraduate degree
f	24	Trentin (local dialect)	housewife	high school diploma
f	26	English	tagesmutter/babysitter	high school diploma
f	25	English(bilingual), German, Russian	hotel receptionist	undergraduate degree
m	32	Spanish, English, German	shop owner/sale assistant	high school diploma
m	32	Trentin (local dialect, bilingual), Spanish, English	barista	terza media (8th grade)
f	40	Trentin(local dialect), English, German	factory labourer	professional school (realschule, 11th grade)
m	43	-	factory labourer	professional school (realschule, 11th grade)
f	43	-	housewife	professional school (realschule, 11th grade)
f	31	Trentin (local dialect, bilingual), English	optoelectronic engineer	PhD
m	32	English	desk clerk	high school diploma
f	30	English	nurse	undergraduate degree

21.2. LOCATIONS

Each nursery endeavoured to provide the Experimenter (E.) with a space dedicated to the study so that no one would interrupt the game. Unfortunately, this was not always possible and the study took place in variety of locations, such as:

a) a quiet room

b) the teachers' office

c) the sleeping room

d) a table in a corridor, usually used for drawing

e) the gym

f) a room normally used for artistic activities

g) a small hall

Adults were tested in a quiet room at the experimenter's house. One two-year-old child underwent both the familiarisation and test phases at home with his/her mother and siblings in attendance.

21.3. THE SCRIPT OF THE INDIVIDUAL SESSION

The following script/play had been followed by the experimenter (E.), who learnt each line of the sequence *Introduction-block_4* by heart.

Each manifestation of the script below contained some natural variations depending on each child's response and readiness to proceed to the next section/block. Importantly though, the various uses of the experimental verbs and the order and manner in which they were presented strictly followed the below script.

21.3.1. WARM-UP

21.3.1.1. Introducing the child to the task

E. and the child arrive at the location where the test phase takes place and find out that *Silvia the Mole* is sleeping.

E.: (talking to the child) *Oh, Silvia sta ancora dormendo*

"Oh, Silvia is still sleeping".

E.: (talking to the child about Silvia) *che dormigliona*

"What a spleepyhead."

E: (talking to the child) *Allora, facciamo così: la lasciamo dormire un po' e intanto ti faccio vedere come giocano le sue amiche Peppa ed Emily.*

"Ok, let's do this: we'll let her sleep a bit and in the meantime I'll show you how her friends *Peppa* and *Emily* play".

E.: (talking to the child) *poi, quando si sveglia, tu le spieghi cosa hanno fatto.*

"Afterwards, when she wakes up, you'll explain to her what they have done."

E.: (talking to the child) *hai voglia di aiutarmi a spiegare alla Silvia cosa fanno le sue amiche?*

"Would you like to help me explain to *Silvia* what her friends do?"

21.3.1.2. Naming the toys

If the child says yes, E. thanks the child and then introduces him/her to Peppa and Emily. E. also makes sure that s/he can pronounce their names. If the child cannot or does not want to pronounce their names, E. and the child negotiate a new name for either or both toys.

Afterwards, E. hides both toys behind his back and then displays one toy at a time while asking the child "who is she?"

Once the child has answered appropriately twice in a row for each toy, E. and the child casually play with them for 1 or 2 minutes.

Afterwards, E. hides both toys behind his back and a further naming trial starts. E. shows the child one toy at a time and asks her/him "is this Peppa/Emily?" while holding either toy (i.e. he asks *is this Peppa?* while holding either Peppa or Emily). Once the child has answered appropriately (*yes, it's Peppa* or *no, it's Emily*, depending on the circumstance) three times in a row (two target answers are negative - *no, its's Emily* – and one is positive – *yes, it's Peppa*), the game moves on to the training phase (training verbs) and then to the experimental phase (familiar verb and nonce verb).

21.3.2. TRAINING AND EXPERIMENTAL PHASE (i.e. Presenting the each verb and eliciting productivity)

Each verb is presented through the sequence *Introduction-block_3* or *Introduction-block_4*. In the example below, the familiar verb *lanciare* "to throw" appears in the "V,A" construction, has Emily as agent and Peppa as Patient[6].

21.3.2.1. Introduction (E: 5 inf; C: 1 inf).

> E.: (talking to the child) *Adesso ti faccio vedere una cosa che si kiama* ***LANCI-ARE***.
>
> "Now, I'll show you something that is called **THROW-INF**."

[6] On each heading of "Introduction-block 4", the reader shall find the total number of imperative (imp) and infinitive (inf) forms of the target verb that were used in each block. The letter C. and the letter E. indicate the child and the Experimenter, respectively. Thus, "E: 5 imp + 1 inf" indicates that E. utters 5 imperative forms of the target verb and 1 infinitive form of the same verb. Similarly, "C: 1 inf" indicates that the child utters one infinitive form of the verb.

E. (talking to the child) *Sai dire **LANCI-ARE?***

"Can you say **THROW-INF**?"

All children were able to correctly pronounce the infinitive form of all verbs.

E: (talking to the child) *Bravo/a! ... adesso ti faccio vedere come si fa a **LANCI-ARE**.*

"Very good, now I'll show you how to **THROW-INF**".

E.: (talking to the child) *Allora ... guarda cosa la EMILY fa alla PEPPA.*

"Ok then, look at what EMILY is going to do to PEPPA."

E.: (talking to the child) *Ti faccio vedere cosa LA EMILY fa ALLA PEPPA.*

"I'll show you what EMILY is going to do to PEPPA."

E. makes Emily act upon Peppa.

E.: (talking to the child) *Hai visto cosa la EMILY ha fatto alla PEPPA?*

"Have you seen what EMILY did to PEPPA?"

E. makes Emily act upon Peppa twice.

E.: (talking to the child) *Vuoi farglielo fare tu, alla EMILY?*

"Do YOU want to make EMILY do that?"

The child makes Emily act upon Peppa as many times as s/he wants. While s/he is making Emily act upon Peppa, E. takes **two** occasions in which the child is clearly paying attention to him to say:

E.: *Ma quanto bello é **LANCI-ARE**?*

"But how cool is it to **THROW-INF**?"

E.: (talking to the child) *Ma ti piace **LANCI-ARE**?*

"Do you like to **THROW-INF**?"

21.3.2.2. Block 1 (E: 4 imp + 2 inf)

E.: (talking to Emily) *Brava, Emily!*

"Well done, Emily!"

E.: (talking to the child) *Posso farglielo fare io, adesso?*

"Can **I** make her do that?"

If the child says no, E. lets the child play for a further minute or so and then asks the above question again. Once the child is happy for E. to have a turn with the agent (i.e. Emily), the game proceeds as follows:

E.: (talking to Emily) *Brava, ancora Emily!*

"Well done, another time, Emily!"

E.: (talking to Emily) *Dai, sù!* ***LANCI-A, EMILY!***

"Com'on, **THROW-IMP.2.SG, EMILY**!"

E. Tries to get EMILY close to PEPPA, but EMILY refuses to act upon Peppa.

E.: (talking to Emily) *Cosa c'é adesso, EMILY?*

"What's wrong now, EMILY?"

E: (talking to Emily) *Non vuoi* ***LANCI-ARE?***

"Don't you want to **THROW-INF**?"

Emily, voiced and moved by E., says "no".

E.: (talking to Emily) *Dai, sù!* ***LANCI-A, EMILY!***

"Com'on, **THROW-IMP.2.SG, EMILY**!"

E: (talking to Emily) **LANCIA, EMILY!**

"THROW-IMP.2.SG, EMILY!"

EMILY: (talking to E.) *ok, va bene, ma solo perché c'è il/la CHILD'S NAME.*

"Ok, fine, but only because CHILD'S NAME is here."

E. makes Emily act upon Peppa.

E.: (talking to Emily) *Brava EMILY, vedi che se vuoi sai **LANCIARE**!*

"Well done, EMILY, you see that you are able to **THROW-INF**, if you want to!"

E.: (talking to Emily) *Ancora, dai! **LANCIA, EMILY!***

"Com'on, again! **THROW-IMP.2.SG, EMILY**!"

E. makes Emily act upon Peppa.

21.3.2.3. Block 2 (E: 3 imp + 1 inf)

E.: (talking to the child) *Vuoi farglielo fare tu alla EMILY* ?

"Do YOU want make EMILY do that?"

In the above line, E. basically invites the child to make the toys perform the action, and therefore the child is made to make Emily act upon Peppa for a minimum of three times.

Afterwards, E. says to the child:

E.: (talking to the child) *Ok, adesso facciamo un nuovo gioco.*

"Ok, let's play a new game, now."

E: (talking to the child) *ogni volta che ordino alla EMILY di **LANCI-ARE**, glielo fai fare tu. Va bene?*

"Any time that I order Emily to **THROW-INF**, YOU make her do that, is that ok?"

At this point, E. orders Emily to perform the action by uttering the imperative stimulus three times. Each time, the child makes Emily act upon Peppa, as shown below.

E: (talking to Emily) ***LANCI-A, EMILY!***

"**THROW-IMP.2.SG, EMILY**!"

The child makes Emily act upon Peppa.

E: (talking to Emily) ***LANCI-A, EMILY!***

"**THROW-IMP.2.SG, EMILY**!"

The child makes Emily act upon Peppa.

E: (talking to Emily) ***LANCI-A, EMILY!***

"**THROW-IMP.2.SG, EMILY**!"

The child makes Emily act upon Peppa.

21.3.2.4. Block 3 (E: 4/5 imp + 1 inf; C: 5 or more imp)

E.: (talking to the child) *Ma che bravo/a che sei!*

"How clever you are!"

E.: (talking to the child) *Mi lasci farlo a me adesso, così divento bravo come te?*

"Can **I** do it now, so that I become as good as you?"

The child hands Emily over to E. and twice the latter orders Emily to act upon the patient (i.e. Peppa), but she refuses both times, as shown below.

E: (talking to Emily) ***LANCI-A, EMILY!***

"**THROW-IMP.2.SG, EMILY**!"

Emily, voiced and moved by E., refuses to act upon Peppa and says "no".

E: (talking to Emily) ***LANCI-A, EMILY!***

"**THROW-IMP.2.SG, EMILY**!"

Emily, voiced and moved by E., refuses to act upon Peppa and says no.

E. then asks the child to persuade Emily to act upon Peppa, as shown below.

E.: (talking to the child) *Mammamia, la Emily non mi ubbidisce. Prova a dirglielo tu di **LANCI-ARE**.*

"Goodness me, Emily is not obeying me. You try telling her to **THROW-INF**."

E.: (talking to the child) *Magari a te ti ascolta.*

"Maybe, she'll listen to you."

E.: dille ***LANCI-A, EMILY! LANCI-A EMILY!***

"Tell her: **THROW-IMP.2.SG, EMILY**! **THROW-IMP.2.SG, EMILY**!"

At this point, the child orders Emily to act upon Peppa five times; each time E. makes Emily act upon Peppa. In the middle of the below sequence, a further failed attempt from E. (identical to previous Emily's refusals to act) is inserted, if necessary (e.g. if the child stops ordering Emily to act upon Patient).

CHILD: (talking to Emily) ***LANCI-A, EMILY!***

"THROW-IMP.2.SG, EMILY!"

E. makes Emily act upon Patient.

CHILD: (talking to Emily) ***LANCI-A, EMILY!***

"THROW-IMP.2.SG, EMILY!"

E. makes Emily act upon Patient.

CHILD: (talking to Emily) ***LANCI-A, EMILY!***

"THROW-IMP.2.SG, EMILY!"

E. makes Emily act upon Patient.

CHILD: (talking to Emily) ***LANCI-A, EMILY!***

"THROW-IMP.2.SG, EMILY!"

E. makes Emily act upon Patient.

CHILD: (talking to Emily) ***LANCI-A, EMILY!***

"THROW-IMP.2.SG, EMILY!"

E. makes Emily act upon Patient.

21.3.2.5. Elicitation (first attempt)

Silvia wakes up and E. draws the child's attention to this.

E.: (talking to the child) *Oh, guarda, Silvia si é svegliata!*

"Oh, look, Silvia woke up."

Silvia approaches and greets the child and then asks him/her:

SILVIA: (talking to the child) *CHILD'S NAME, mi dici cosa ha fatto la Emily?*

"CHILD'S NAME, can you tell what Emily did?"

SILVIA: (talking to the child) *che cosa ha fatto la Emily?"*

"What did Emily do?"

If the child answers using the target verb (i.e. *lanciare* "to throw") in the target constructions – i.e. *(Emily)+ha<lanciato<Peppa* "(Emily)+has<thrown<Peppa" or *(Emily)+l'<ha<lanciata* "(Emily)+her(clitic)<has<thrown", the games move on to the "in-between verbs" phase (section 21.3.2.8) and then to another verb.

However, if this does not happen, the games moves on differently, depending on whether the answer (or non-answer) pertains to the training verbs or the verbs in the experimental phase (i.e. familiar verb and nonce verb).

If, **during training** (*lavare* "to wash" and *pettinare* "to comb"), the child:

a) uses the target verb, but does not provide the target answer, or
b) uses a verb different from the target one used in the game, or
c) stays silent,

s/he is helped to produce the target answer. If needed, E. suggests the answer to the child who is encouraged to repeat it to Silvia several times. Afterwards the game moves on to the "in-between verb phase" (21.3.2.8).

If, during the **test phase** (*lanciare* "to throw" and *bodare* "to nonce.verb"), the child

a) provides a syntactically and morphologically unproductive answer, or
b) uses a verb different from the target one used in the game, or
c) stays silent,

a further block of stimuli (*block_4*) is in order.

21.3.2.6. Block 4 (optional; E: 5 imp + 2 inf; C: 3 or more imp)

E. attempts to provide an answer to Silvia but she falls asleep before he can finish his sentence. Thus, E. suggests that the child and he go on playing a bit more, while Silvia sleeps. All children were thrilled to do so.

E.: (talking to Silvia) *Silvia, è successo che +//.*

"Silvia, it happened that +//."

Silvia the Mole falls asleep.

E.: (talking to the child) *Ma hai visto?!*

"Have you seen!?"

E.: (talking to the child about Silvia) *Questa si è addormentata ancora!*

"She fell asleep, again!"

E.: (talking to the child) *Allora, noi andiamo avanti e quando si sveglia tu riprovi a spiegarle cosa ha fatto la Emily. Va bene?*

"Well, we'll carry on playing and when she wakes up, you try explaining to her what Emily did, ok?"

All children confirmed that they were happy to continue.

E.: (talking to the child) *grazie, sai, che mi aiuti, perché la Silvia è proprio una dormigliona.*

"Thanks. You know, you are really helping me, because Silvia is a real sleepy-head."

E.: (talking to Emily) *Emily, hai ancora voglia di **LANCI-ARE**?*

"Emily, do you still want to **THROW-INF**?"

Emily, voiced and moved by E., enthusiastically agrees.

E.: (talking to Emily) ***LANCIA, EMILY!***

"THROW-IMP.2.SG, EMILY!"

E. makes Emily act upon Patient.

E.: (talking to Emily) ***LANCIA, EMILY!***

"THROW-IMP.2.SG, EMILY!"

E. makes Emily act upon Patient.

At this point, Emily starts ignoring E., who asks for the child's help, as shown below.

E.: (talking to Emily) ***LANCIA, EMILY!***

"THROW-IMP.2.SG, EMILY!"

Emily, voiced and moved by E., refuses to act upon Peppa and says "no".

E.: (talking to the child) *Mammamia, la Emily non mi ubbidisce.*

"Goodness me, Emily is not obeying me."

E.: (talking to the child) *Prova a dirglielo tu di* ***LANCI-ARE.***

"YOU try telling her to **THROW-INF**."

E.: (talking to the child) *Dille,* ***LANCI-A, EMILY! LANCI-A, EMILY!***

"Tell her **THROW-IMP.2.SG, EMILY**! **THROW-IMP.2.SG, EMILY**!"

At this point the child orders Emily to act upon Peppa three times. Each time E. makes the toys perform the target action, as shown below.

CHILD: (talking to Emily) ***LANCI-A, EMILY!***

"THROW-IMP.2.SG, EMILY!"

E. makes Emily act upon Peppa.

CHILD: (talking to Emily) ***LANCI-A, EMILY!***

"THROW-IMP.2.SG, EMILY!"

E. makes Emily act upon Peppa.

CHILD: (talking to Emily) ***LANCI-A, EMILY!***

"THROW-IMP.2.SG, EMILY!"

E. makes Emily act upon Peppa.

21.3.2.7. Elicitation (second attempt; optional)

Silvia wakes up and E. draws the child’s attention to this.

E.: (talking to the child) *Oh, guarda, Silvia si é svegliata!*

“Oh, look, Silvia woke up.”

Silvia approaches and greets the child and then asks him/her:

SILVIA: (talking to the child) *CHILD’S NAME, mi dici cosa ha fatto la Emily?*

“CHILD’S NAME, can you tell me what Emily did?”

SILVIA: (talking to the child) *che cosa ha fatto la Emily?”*

“What did Emily do?”

If the child provides an answer, E. compliments, thanks and high-fives the child. The game then moves on to the “in-between verbs phase”.

If the child does not provide an answer (or uses a verb other than the target verb), E. notes the answer as unresponsive and moves on to another verb trial. When this happens, Silvia wakes up and E. attempts to explain what happened. However, the mole falls asleep before he can finish his sentence and then the game moves on to the “in-between verbs phase”.

E.: (talking to Silvia) *Silvia, è successo che +//.*

“Silvia, it happened that +//.”

Silvia falls asleep.

21.3.2.8. In-Between Verbs

Silvia falls asleep.

E.: (talking to the child) *Ma hai visto?!*

“Have you seen!?”

E.: (talking to the child) *Questa si è addormentata ancora!*

“She fell asleep, again!”

E.: (talking to the child) *Vabbè, dai, lasciamola dormire.*

"Oh well, let's let her sleep!"

E.: (talking to the child) *Hai voglia di fare un altro gioco e poi quando Silvia si sveglia, tu le spieghi ancora cosa succede?*

"Do you want to play another game and then, when Silvia wakes up, you explain to her again what happened?"

If the child says yes, E. adds the following:

E.: (talking to the child) *Ma sei proprio gentile, sai.*

"You're really kind, you know."

E.: (talking to the child) *Grazie.*

"Thank you."

The games moves on to another verb-action pair and re-starts from the Introduction (21.3.2.1).

21.4. SECTION 11.4: The adapted version of Cianchetti and Sannio Facello's (2010) *test di valutazione del linguaggio* (tvl)

In order to assess children's vocabulary, Cianchetti and Sannio Fancello's (2010) *test di valutazione del linguaggio* "test for the assessment of language" (**tvl**) has been adapted to suit this experiment. The tvl is divided into four sections: (1) comprehension of phrases and words, (2) sentence repetition, (3) production and (4) elicited spontaneous production. The adapted version attended by children included some parts of (1) and the whole production section (3). Because some items repeated themselves across (1) and (3), production (3) and comprehension (1) were administered in separate sessions that were at least one day apart. The following paragraphs and figures illustrate the task children had to attend.

21.4.1 Comprehension (day 1; 5 to 10 minutes)

The comprehension part uses the following sections of the tvl: 1.1a (body parts), 1.2 (pictures of objects), 1.3 (colours) and 1.5 (adjectives).

Body Parts

In section 1.1a of the tvl the child participant is shown a drawing of a child (fig. 21.1) and is asked by E. to point to ten different body parts in the drawing (*show me where the child's forehead is*, *show me where the child's shoulders are*): nose, mouth, legs, hands, feet, arms, cheeks, front, shoulders and knees.

Figure 21.1: tvl, comprehension (1.1a), from Cianchetti and Sannio Fancello (2010, p 25).

Pictures of objects

In section 1.2, the child is shown five drawings, like the one in fig. 21.2. Each drawing contains six objects arranged into two lines of three. For each drawing, knowledge of four of the six objects' names is tested by asking the child to point to the named object (*dov'è il treno?* "Where's the train?"). Overall, knowledge of 20 names is investigated: ice-cream, apple, television, train, broom, bow, bread-roll, flag, penguin, arrow, lake, telescope, door-handle, pine-cone, grain-spike, fence, parachute, locker, funnel and astronaut.

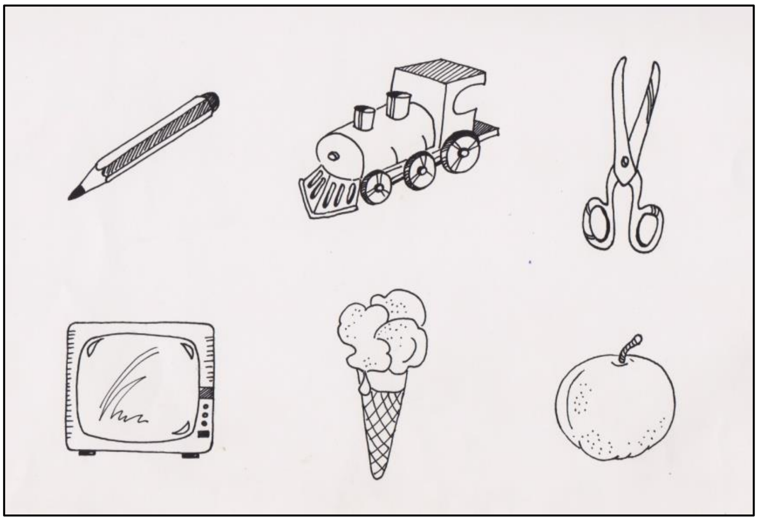


Figure 21.2: comprehension (1.2), from Cianchetti and Sannio Fancello (2010, p. 31).

Colours

In 1.3, the child is shown a page divided into two columns, with each column showing five different coloured rectangles (fig. 21.3). E. asks the child to point to a named colour (*mi fai vedere qual'è il rosso* "Would you show me which one is the red one?"). Overall, knowledge of 10 colours is tested.

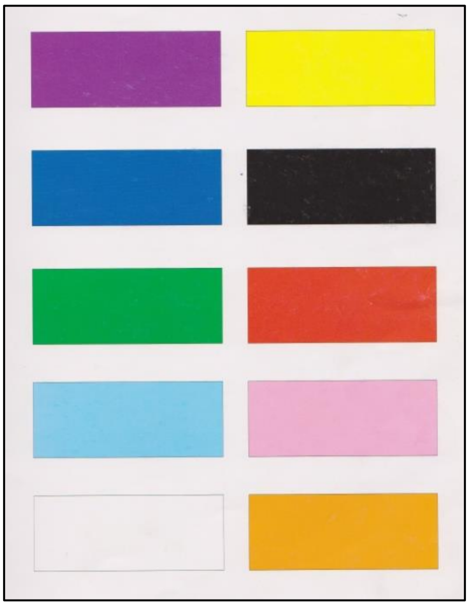


Figure 21.3: comprehension (1.3), from Cianchetti and Sannio Fancello (2010, p. 37).

Adjectives

In section 1.5, the child is tested with 15 noun-adjective sequences, and 5 noun-adjective-adjective sequences.

The first eight items tap into children's ability to tell apart two instantiations of the same noun based on the adjective with which it is combined. For instance, the child is shown a pair of televisions, one bigger than the other (fig. 21.4), E. then asks the child *dov'è il televisore grande?* "Where is the big television?" The child must then point to the appropriate object on the basis of the adjective that modifies it (in this case the one on the right).

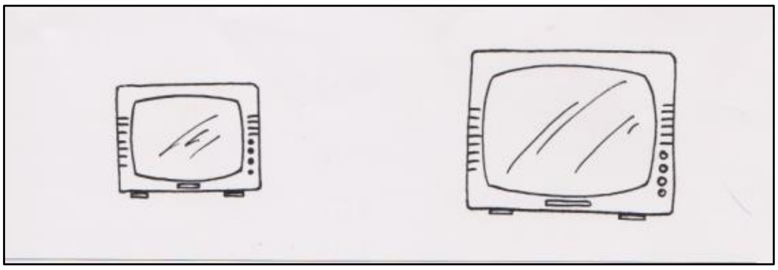


Figure 21.4: comprehension (1.4), 1 of 4. From Cianchetti and Sannio Fancello (2010, p. 42).

A further four items expose the child to noun-adjective sequences, but the difficulty of the task increases in two ways. Firstly, the adjectives used are conceptually more demanding and secondly, the child must now choose between a greater number (two to six) of options. For instance, in fig. 21.5, E. asks the child *dov'è la bottiglia vuota?* "Where's the empty bottle?" and the child has to point towards the right object (the second, and not the fifth, from the left).

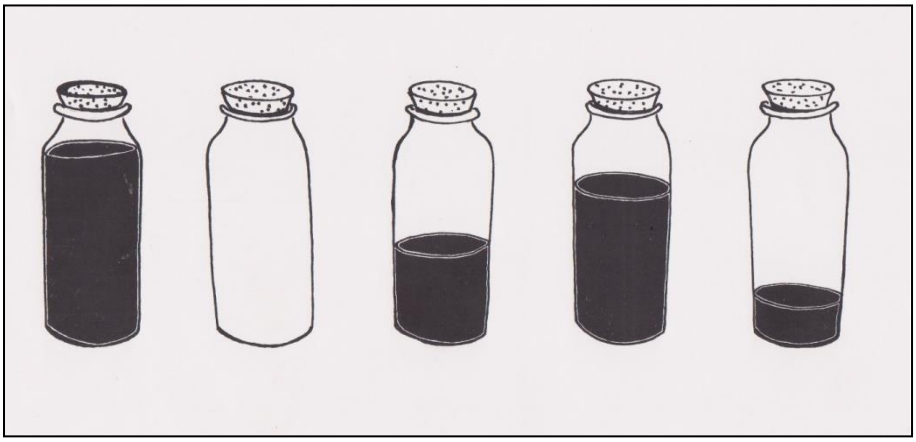


Figure 21.5: comprehension (1.4), 2 of 4. From Cianchetti and Sannio Fancello (2010, p. 42).

A further three trials ask the child to choose amongst multiple candidates that present more complex adjectives. Thus, the child is shown fig. 21.6 and is asked *dov'è la scatola con meno palline?* "Where's the box with fewer balls?"

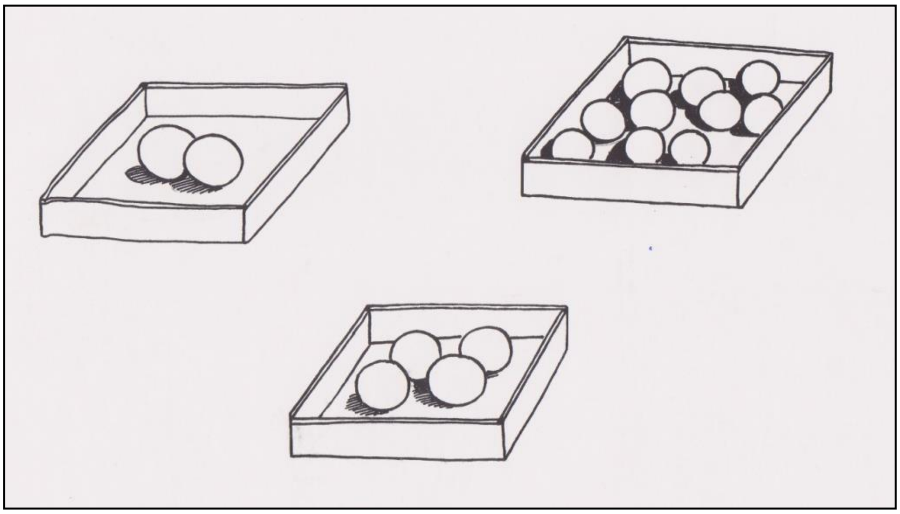


Figure 21.6: comprehension (1.4), 3 of 4. From Cianchetti and Sannio Fancello (2010, p. 42).

The final 5 items are noun-adjective-adjective sequences: the child is shown a set of pictures (e.g. fig 21.7) and E. asks him/her *dov'è il cane grande bianco?* "Where's the big white dog?". The child must point at the correct target on the basis of both adjectives, as attending only one of them might lead him/her to point to either a big dog that is not white or a white dog that is not big.

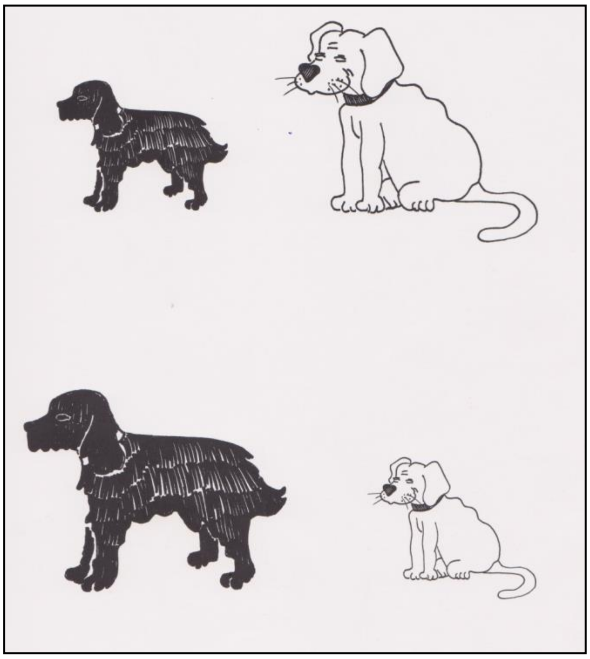


Figure 21.7: comprehension (1.4), 4 of 4. From Cianchetti and Sannio Fancello (2010, p. 51).

21.4.2. Production (day 2; duration: about 5 minutes)

Naming real objects

Firstly, the child is put in front of three sets of real objects. The child encounters one set at a time with each set comprising six toy objects organised in two rows, as in fig. 21.8. E. asks the child to name the target object by asking *come si chiama questo?* “What’s this called?” Overall, the production of ten nouns is elicited.

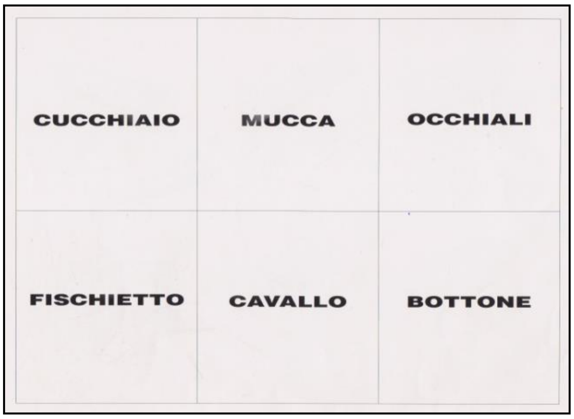


Figure 21.8: production (3), from Cianchetti and Sannio Fancello (2010, p. 27). The figure illustrates how objects have to be positioned in front of the child. In the figure (from top to bottom and from left to right), the objects the children saw were: a spoon (*cucchiaio)*, a cow (*mucca*) a pair of glasses (*occhiali*), a whistle (*fischietto*), a horse (*cavallo*) and a button (*bottone*).

Naming body parts and pictures of objects

Afterwards, the same items used to test comprehension in the “body parts” (fig. 21.1) and “pictures of objects” (fig. 21.2) sections are now used to test production using the following trigger: *come si chiama questo?* ”What’s this called?”

21.4.3 Coding

The following outlines the method used to score the child’s comprehension and production of the target lexical items.

As for comprehension, when the child points at the right target, the answer is assigned 1 point, when s/he indicates the wrong object the answer is given 0 points.

As for production, when the child answers quite straightforwardly with the appropriate name, the answer is given 1 point. If the child remains silent, E.

suggests the first syllable of the name. If the child is then capable of finishing the word started by E., the answer is assigned 0.5 points. The same number of points (0.5) is assigned whenever the child names the object correctly, but the word contains two or more phonological errors, which do not compromise communication. If none of the above three conditions realised, the answer (or non-answer) is assigned 0 points.

21.5. SECTION 11.6.2: training, familiar and nonce verbs in the main and preliminary studies

As for children who participated in the main study, section 11.2.2 illustrated that *lavare* "to wash" and *pettinare* "to comb" were chosen as training verbs, whereas *lanciare* "to throw" was chosen as control-familiar verb for the test phase.

However, children who took part to the preliminary study had slightly different known, real verbs, while the nonce verb (*bodare*) was the same.

Two and four year-olds in the preliminary study had *pettinare* "to comb" as the familiar verb, which is one of the training verbs in the main study. The training verbs in the preliminary study were *lavare* "to wash", which is also a training verb in the main study, and *coccolare* "to cuddle".

With *coccolare* "to cuddle", the agent approaches the patient and then hugs and caresses her. In Italian the action of *cuddling* is more frequently expressed with the construction *fare le coccole* "make/give some cuddles" than with the transitive form of the verb. This more frequent construction pre-empted the use of the transitive form of the verb and children never used the (elicited) constructions. Hence, *coccolare* "to cuddle" was substituted with *lanciare* "to throw".

Three-year-olds in the preliminary study had *lavare* "to wash" and *lanciare* "to throw" as training verbs and *pettinare* "to comb" as the familiar (control) verb.

Lanciare "to throw" turned out to be more salient than *pettinare* "to comb". The new action-verb pair was greatly enjoyed by both females and males, whereas *to comb* was not as successful with boys, who often wanted to move on to another game when *combing* was involved. Hence, *lanciare* "to throw" was selected as the familiar (control) verb and *pettinare* “to comb” was “downgraded” to a training verb in the main study

Since four out of the five three-year-old subjects in the preliminary study were able to successfully explain to *Silvia the Mole* what the agent had done to the patient with respect to the target verb *lanciare* “to throw”, their answers have been considered for the analysis. That is, they answered the Mole’s questions using the target verb (e.g. (*Emily)* + *ha<lanciato<Peppa* “Emily has thrown Peppa”) and did so without needing any help.

Furthermore, since all children in the preliminary study had the same nonce verb (*bodare*) as children in the main study, their answers with the test-nonce verb have also been considered for analysis. Table 21.2 summarises how data (children’s answers) have been obtained.

Table 21.2: The data collected during preliminary and main study. The answers relating to the specific verbs that have been used for the analysis are in CAPITALS.

study	participants	no. of participants	training verb 1	training verb 2	familiar verb	nonce verb
preliminary	2-year-olds	3	*lavare* ‘to wash’	*coccolare* ‘to cuddle’	*Pettinare* ‘to comb’	BODARE
preliminary	3-year-olds	4	*lavare* ‘to wash’	*LANCIARE* ‘to throw’	*pettinare* ‘to comb’	BODARE
preliminary	4-year-olds	5	*lavare* ‘to wash’	*coccolare* ‘to cuddle’	*pettinare* ‘to comb’	BODARE
main	all participants	55	*lavare* ‘to wash’	*pettinare* ‘to comb’	*LANCIARE* ‘to throw’	BODARE

21.6. CHAPTER 14: A (non-exhaustive) introduction to Classification Regression Trees (CART)

The method of Classification (categorical outcome variable) and Regression (continuous outcome variable) trees (CARTs) is also called recursive partitioning. In this method, the space represented by the data is recursively partitioned into subspaces, which group similar values of the response variable in a binary fashion, based on the values associated with the predictors.

Fig. 21.9-21.13 are based on fig. 1 of Berk (2006, p. 266) and represent a graphical representation of how recursive partitioning works. Following Berk (2006), let us assume a dichotomous outcome variable *Y*, whose values are either A or B and two continuous (or ordinal) predictors *X* and *Z*. The dataset can be thought of as a square within which the values of *Y* collocate (fig. 21.9).

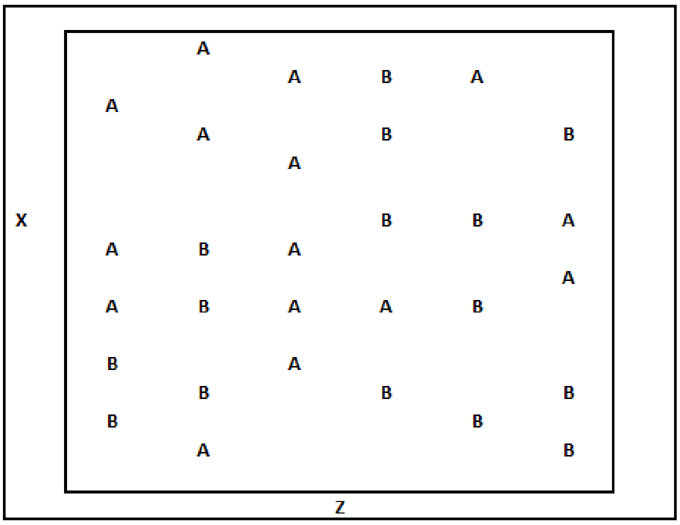


Figure 21.9: Classification tree: 1 of 5.

The dataset is firstly divided into two zones, which group similar values of *Y*. In figure 21.10, the dataset is firstly vertically divided on the basis of the values taken by predictor *Z*: values ≤ 13 are assigned on the left zone, the ones higher than 13 end up in the right zone. On the left hand-side there are more As, on the right hand-side there are more Bs.

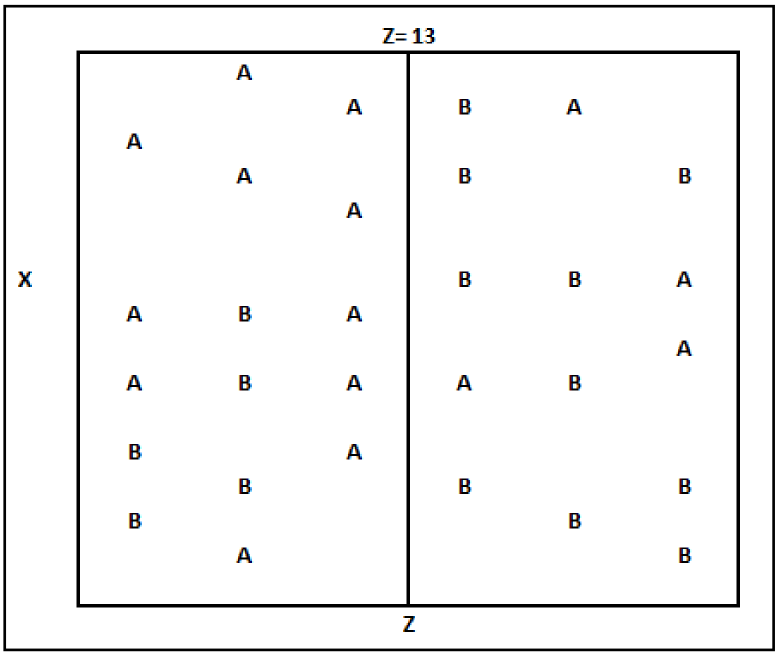


Figure 21.10: Classification trees: 2 of 5.

Each subspace is then partitioned again; this time on the basis of the values taken by *X*. In figure 21.11, the left space is further partitioned: values higher than 33 are assigned to the top (left) zone, in which all *Y*s are A.

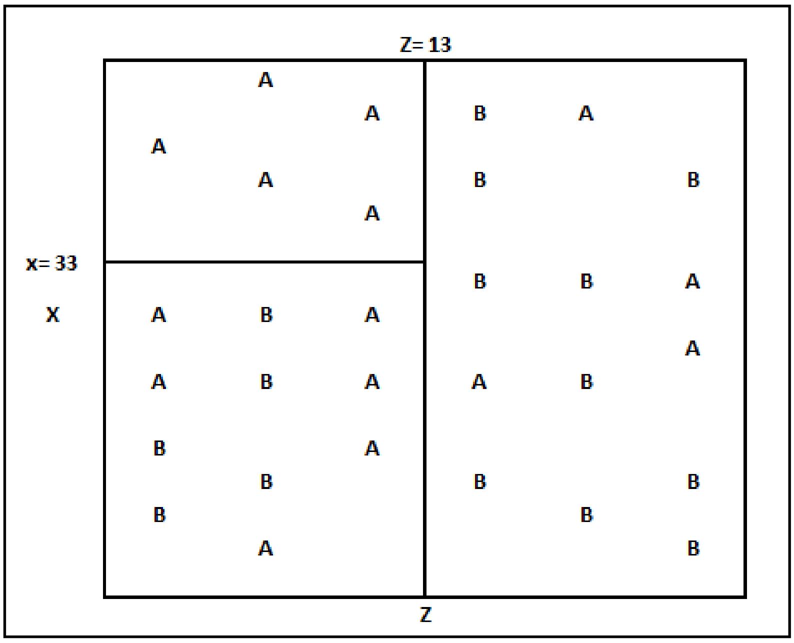


Figure 21.11: Classification trees: 3 of 5.

In figure 21.12, the final partitioning splits the right zone into a top space, where values of X are higher than 20, and a bottom space, where $X \leq 20$. In the latter subzone all *Y*s are B. Such a partitioning is graphically represented with inverted trees (figure 21.13).

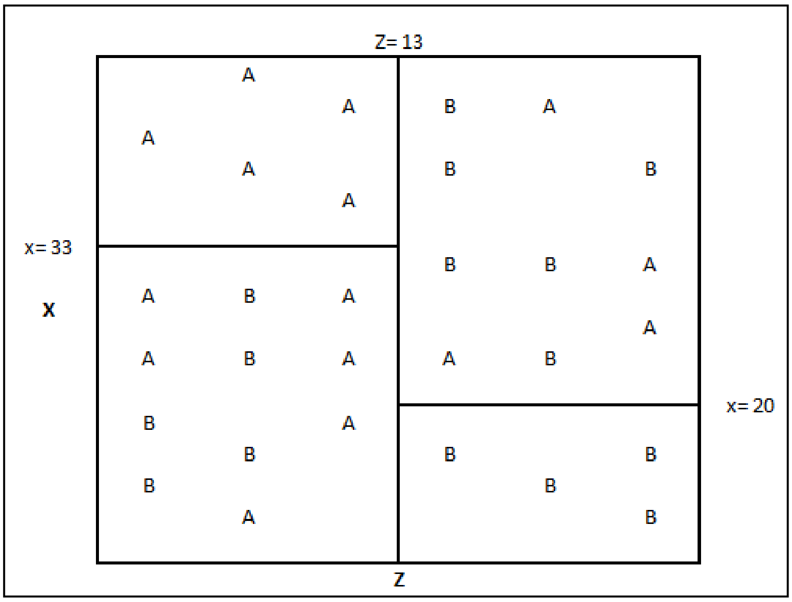


Figure 21.12: Classification trees: 4 of 5.

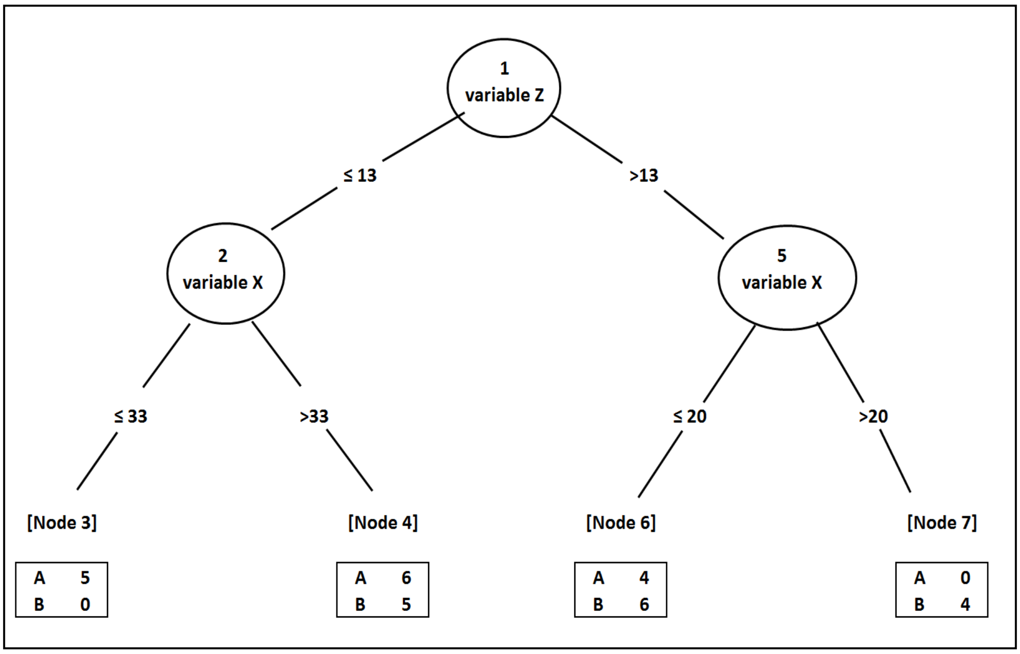


Figure 21.13: Classification trees: 5 of 5.

In recursive partitioning methods, the splitting points (i.e. node creation) are selected following the principle of impurity reduction (Strobl et al., 2009; Berk, 2006). Looking back at figures 21.9-21.12, it can be noted that spaces are created that progressively present more uniformity of values of the response variable, i.e. the daughter nodes contain purer values than their parents. A pure node contains only one value of a response variable (node 3 and 7 in fig. 21.13). Such impurity reduction is measured

> “by the difference between the impurity in the parent node and the average impurity in the two daughters”
>
> (Strobl et a., 2009, p. 5).

Different CART methods adopt different entropy measures to calculate impurity, such as *Gini Index* and *Shannon Entropy* (see Strobl et al., 2009; Berk, 2006). Strobl et al. (2009) note that the impurity reduction can be thought of as a way to measure the association between predictor and outcome variable.

As did Strobl et al. (2009) and Boyd and Goldberg (2012), I adopted Hothorn et al.’s (2006) unbiased classification tree algorithm, which uses the p-values of the association tests between predictors and DV to select the appropriate splitting-node. Indulging in some simplification, the predictor that has the smallest p-value (strongest association with the outcome variable) is selected first. When the selected predictor has more than two categories (e.g. *d, f, g, h*) (or it is continuous) the cutpoint for node splitting (e.g. [*d, f*] / [*g, h*] rather than [*d*] / [*f, g, h*]) is selected on the basis of which binary split yields the purest subsets (the purest node). Thus, the method forces the partitioning that better groups together similar values of the response variable (A or B) on the basis of the values attested in the chosen predictor (in this case, ≤13 as value of variable *Z*; node 1 in fig. 21.13). Once a node is created, all further splits are node-internal.

This process of partitioning can be carried on until there is nothing left to split. I use the R package *partykit* (https://cran.r-project.org/web/packages/partykit/index.html), which allows the selection of a specific p-value as stopping criterion of tree-growth.

As a final remark, it is important to bear in mind that a predictor can be selected multiple times. So, if *X* is selected for the first node and the following nodes are determined by *Z* and *W*, then *X* can still be selected for a further node.