Different speakers, different grammars: Individual differences in native language attainment

Ewa Dąbrowska
Department of Humanities
School of Arts and Social Sciences
Northumbria University
Newcastle-upon-Tyne
NE1 8ST
UK
ewa.dabrowska@northumbria.ac.uk

This article reviews several recent studies suggesting that – contrary to a widespread belief – adult monolingual native speakers of the same language do not share the same mental grammar. The studies examined various aspects of linguistic knowledge, including inflectional morphology, passives, quantifiers, and more complex constructions with subordinate clauses. The findings suggest that in some cases, language learners attend to different cues in the input and end up with different grammars; in others, some speakers extract only fairly specific, 'local' generalizations which apply to particular subclasses of items while others acquire more abstract rules which apply 'across the board'. At least some of these differences are education-related: more educated speakers appear to acquire more general rules, possibly as a result of more varied linguistic experience. These findings have interesting consequences for research on bilingualism, particularly for research on ultimate attainment in second language acquisition, as well as important methodological implications for all language sciences.

Keywords: individual differences, ultimate attainment, first language acquisition, second language acquisition, academic achievement, passive, quantifier, subordination, inflectional morphology

"… children in the same linguistic community all learn the same grammar." (Crain and Lillo-Martin 1999: 9)

"… children are exposed to different samples of utterances but converge on the same grammar." (Seidenberg 1997: 1600)

"Children of the same speech community reliably learn the same grammar." (Nowak et al. 2001: 114)

"L1A is uniformly successful, with all normal children attaining full competence, whereas in L2A there are various outcomes." (Birdsong 2004: 83)

"The set of utterances to which any child acquiring a language is exposed is equally compatible with many distinct descriptions. And yet children converge to a remarkable degree on a common grammar, with agreement on indefinitely many sentences that are novel. Mainly for this reason, Chomsky proposed that the child brings prior biases to the task." (Lidz and Williams 2009: 177)

As these quotations illustrate, many linguists believe that all first language learners attain more or less the same grammar at the end of the acquisition process, which is usually assumed to be complete by about age five or even earlier. While this view is most strongly associated with generativists, it is also widely accepted by linguists and cognitive scientists with other theoretical orientations (cf. the Seidenberg quote). It is, of course, well known

that there are large individual differences in vocabulary size and knowledge of highly literary constructions (*Little did I know that…; Had she known this….*), and that this knowledge continues to develop well into adulthood. However, when it comes to basic grammar – speakers' representations of the general morphological and syntactic patterns found in the language – all adult speakers are thought to share the same system: in fact, this claim is sometimes used as evidence for other, more controversial claims (e.g. the existence of an innate universal grammar, as in the last quotation). Conversely, in L2 research, the fact that second language learners do not converge is sometimes used to argue that L2 learning is "fundamentally different" from first language acquisition (Bley-Vroman 1990, 2009).

Assertions about convergence in L1 acquisition are rarely justified, presumably because they are regarded as self-evident: we can understand one another, so we must have the same grammar. But this is clearly a *non sequitur*: sharing the same grammar is not essential for successful communication. (Many L2 learners communicate very efficiently with native speakers, for example, but patently do not share the same grammar.) Authors who attempt to substantiate the convergence claim sometimes appeal to the (presumed) agreement about the grammaticality or ungrammaticality of "indefinitely many sentences that are novel" – typically without citing any relevant evidence (cf. Lidz and Williams). Indeed, all speakers of English presumably agree that a simple sentence like *The cat sat on the mat* is well-formed while **Cat the mat sat the on* is not. However, as soon as we start eliciting information about theoretically more interesting structures, there is widespread

disagreement (Ross 1979; see also Schütze 1996 and Cowart 1997 for insightful discussions).[1]

In this paper, I argue that the claim that all learners converge on the same grammar is a myth. I review a series of studies examining aspects of native speakers' knowledge of inflectional morphology (the Polish genitive and dative inflections) and syntax (English passives, the universal quantifier *every*, and several types of sentences containing subordinate clauses). As we will see, some of these constructions are not fully mastered by all native speakers; in other cases, "the same" knowledge is represented differently by different speakers. Apart from one experiment which also involved second language learners (Dąbrowska and Street 2006), the participants in the studies were monolingual adults who were aged 18 or above, had no known learning disability, and had not been exposed to any other language while growing up, except in foreign language lessons at school. However, since the notion of a mature native speaker's knowledge is a benchmark against which learners' performance is typically assessed, the findings are relevant for research on both first and second language acquisition, as well as having general methodological implications for all language sciences.

[1] Schütze and Cowart both stress that grammaticality judgment experiments can provide stable and reliable measures of the grammaticality/ungrammaticality of particular sentence types. However, these are obtained by averaging over individual judgments: in other words, they are group data which mask individual variation.

Polish genitive singular masculine inflection

The first two examples we will consider both involve the Polish case marking system, specifically the genitive masculine (this section) and the dative inflections (the following section). The Polish genitive marking system is complex and quite irregular (see Dąbrowska 2001). There are three main endings: -*i/y* (usually regarded as variants of the same ending), -*a* and -*u*. The choice of ending is partly determined by gender: feminine nouns take -*i/y*, neuter nouns take -*a*, and masculine nouns take either -*a* or -*u*. However, there are no reliable rules determining the choice of the two masculine endings, although there are some broad regularities. Some of these are semantic: for instance, nearly all animate nouns, and a substantial majority of nouns designating body parts and small easily manipulable objects, take -*a*, while nouns designating substances, locations, collections of objects, and abstract concepts usually take -*u*. Others are morphological and phonological: some derivational affixes and stem-final consonants or consonant clusters are associated with -*a*, some with -*u*. However, there are many exceptions to these tendencies, and they are sometimes in conflict: for instance, some abstract nouns, which normally take -*u*, end in palatalised consonants, which strongly favour -*a*.

Thus, it is not clear what the 'correct' generalization(s) would be. A learner could note that most masculine nouns take -*a* and learn one simple rule ("add -*a* to the stem if the noun is masculine") with a large number of exceptions. Alternatively, a learner could note that virtually all animate nouns take -*a*, while about two-thirds of inanimates take -*u*, and learn a more complex rule ("add -*a* if the noun is masculine and animate and -*u* if the noun is masculine and inanimate") with fewer exceptions. Finally, learners could extract more

specific generalizations for narrower semantic classes (small objects, locations, substances, etc.) or subclasses defined in morphological or phonological terms.

So what kinds of generalizations do learners extract from the input? Do they learn relatively general rules with many exceptions, or do they prefer more specific and more reliable generalizations? Dąbrowska (2008a)[2] describes a nonce word inflection experiment designed to determine whether Polish speakers are sensitive to one of the most reliable semantic regularities: that nouns designating small easily manipulable objects typically take *-a*, while most nouns designating substances take *-u*. Adult native speakers of Polish (N=26) were presented with nonce nouns referring to unfamiliar objects and substances in the presentational construction (*jest* … 'here's the …') and asked to use them in a grammatical context requiring the genitive (after the negative expression *nie ma* 'there is no'). For instance, on hearing the cue *jest figon* 'here's the figon', the participant would respond with *nie ma figona* or *nie ma figonu* 'there is no figon'. There was also a control condition in which the referent of the nonce noun was not shown to the participants, although it was clear from the linguistic context that the noun designated an inanimate entity. The nonce words were one to three syllables long and had the phonotactic structure of real Polish words and gender-typical offsets (i.e., all the masculine nouns ended in a 'hard' consonant).

[2] Dąbrowska (2008a) discusses data for four age groups: six-, ten- fourteen-, and eighteen-year-olds. The analysis presented here is based on data for the 20 eighteen-year-olds from the original study, plus six additional participants.

Figure 1 shows individual participants' sensitivity scores, which were computed by adding the number of *-a* responses in the object condition and the number of *-u* responses in the substance condition and converting the figures into percentages. A score of 100% indicates that a participant consistently used *-a* with nouns designating novel objects and *-u* with nouns designating substances, while a score of 50% indicates that a participant was equally likely to use each ending in both conditions, and hence showed no sensitivity to the semantic properties of the noun. As can be seen in the figure, two of the participants appear to be consistently applying the semantic rule, while the remaining 24 are at chance.

Figure 1: Individual sensitivity to the substance/object contrast in masculine nouns

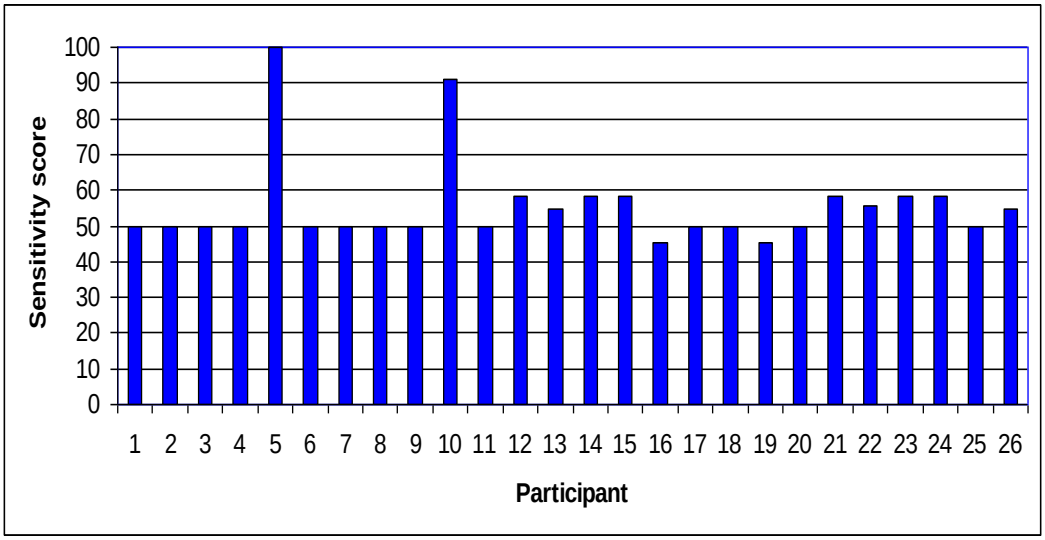
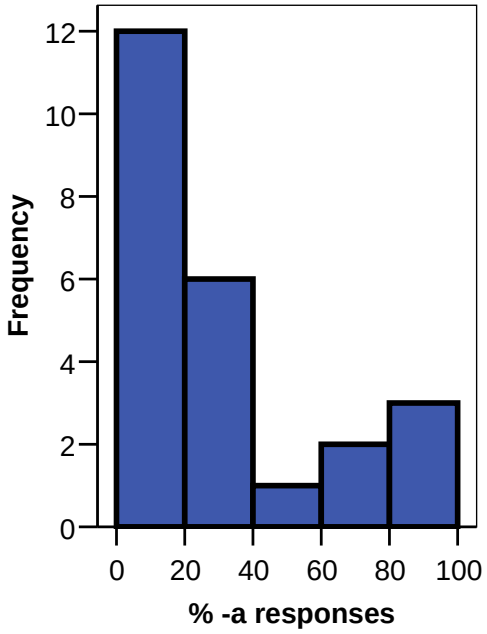


Figure 2: Individual preferences for -*a* or -*u*



The twenty-four participants who were not sensitive to the object/substance contrast used -*a* approximately 31% of the time and -*u* 69% of the time. This closely reflects frequencies of the two endings with inanimate nouns in the real lexicon, and suggests that speakers may be applying them probabilistically. However, an analysis of individual speakers' responses suggests a different picture. Figure 2 presents information about the frequency of individual speakers' use of -*a* as a proportion of all target (i.e. -*a* or -*u*) responses in all three conditions. (The two participants who were sensitive to the object/substance cue were excluded from this analysis.) If speakers simply used the two endings probabilistically, we should have a normal distribution peaking at about 30%, since this is the relative frequency of the -*a* ending. However, the distribution in Figure 2 is clearly bimodal: about half of the participants used -*a* with less than 20% of the nonce

nouns, thus showing a strong preference for -*u;* while a smaller but still sizeable group (about 13%) had an equally strong preference for -*a*.

These results suggest that different speakers extracted different rules. Some used -*a* with all, or nearly all, nouns on the test, in spite of the fact that they had inanimate referents, which suggests that the had one general rule: masculine nouns take -*a*. Other participants consistently used -*u*, which suggests that they had learned two more specific rules: inanimate masculine nouns take -*u*, animate masculine nouns take -*a*. (Dąbrowska and Szczerbiński 2006 have shown that Polish-speaking adults consistently use -*a* with nouns referring to animates.) A small minority appeared to rely on even more specific rules for narrower semantic subclasses of nouns ("masculine nouns designating small easily manipulable objects take -*a*"; "masculine nouns designating substances take -*u*"). Finally, some participants used both endings but did not appear to be sensitive to semantic factors other than animacy. These speakers may be applying the two endings probabilistically or relying on phonological cues.

Polish dative singular inflection

The genitive masculine inflection is in many ways a special case: the various regularities are only partial, and hence it is not clear what the 'correct' generalization would be. It is, therefore, not entirely unexpected that different learners extract different generalizations. But would we find similar differences in more regular linguistic subsystems?

Unlike the genitive, the Polish dative singular inflection is almost completely regular. There are four main endings: -*owi* for masculine nouns, -*u* for neuter nouns, and –*e* or -*i/y* for feminines. The distribution of the feminine endings is determined by phonological properties of the last consonant of the stem. Although the relevant rules are fairly complex (see Bielec 1998, Orzechowska 1998), the distribution of the endings is nevertheless highly predictable. There are also some relatively small classes of nouns with unusual properties: adjectival nouns (which decline like adjectives), masculine nouns ending in -*a* or -*o* (which decline like feminines) and indeclinable nouns (which, as the name suggests, do not decline at all). All of these are systematic exceptions in the sense that they apply in all cases, not just the dative, and are fairly easily identifiable (i.e., the relevant nouns are non-canonical in various ways). Finally, there also some lexical exceptions: about 20 masculine nouns take -*u* rather than -*owi*.

Dąbrowska (2008b) describes a nonce word inflection experiment testing productivity with the dative inflections. Thirty-six adult native speakers of Polish of varying educational backgrounds (from 8 to 20 years in full-time education) were presented with nonce nouns in the nominative and asked to use them in grammatical contexts requiring the dative (after the preposition *dzięki* 'thanks to' or the verb *przyglądać się* 'to look at/watch attentively'). One third of the nonce nouns were masculine, one third were feminine, and one third neuter; each noun's gender could be reliably inferred from its phonological form. Within each gender, half the nouns belonged to densely populated neighbourhoods (i.e., resembled many existing nouns) and half to sparsely populated neighbourhoods. An example of a test item is given in (1) below.

(1) **Zagezia** to duży ptak podobny do strusia. Dzieci przyglądały się _____.

'A **zagezia** is a large bird similar to an ostrich. The children looked at the _____.'

According to usage-based models, productivity is largely a function of type frequency (Bybee 1995, 2001; Dąbrowska and Szczerbiński 2006; Tomasello 2003), and therefore differences in the number of nouns that speakers experienced with each ending should translate into differences in productivity. Since the number of nouns experienced in any particular case form should be strongly correlated with noun vocabulary size, it was hypothesised that individuals with larger vocabularies would be more productive with the endings. To test this hypothesis, the participants were also given a multiple-choice vocabulary test.

Figure 3: Performance on inflection task plotted against vocabulary score

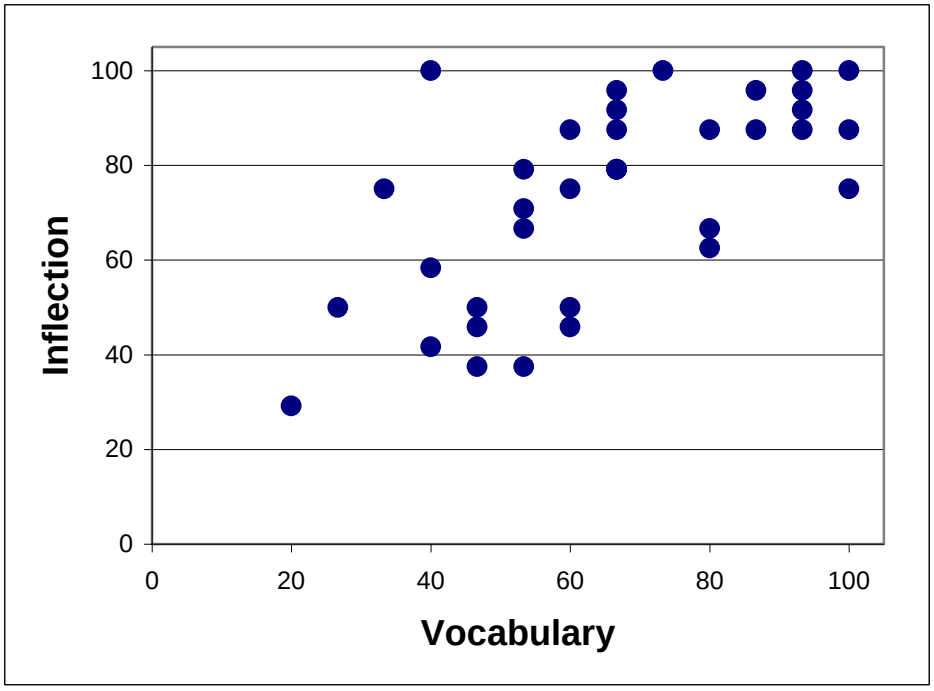


Figure 3 presents performance on the inflection task plotted against vocabulary scores. Individual scores on the inflection task ranged from 29% to 100% (for nouns from low-density neighbourhoods, from 4% to 100%) and were similar in magnitude to differences in vocabulary scores (20%-100%). As predicted, there was a significant correlation between performance on the two tasks ($r = 0.65$, $p < 0.001$). However, there was an even stronger correlation between the inflection task score and the number of years spent in full-time education ($r = 0.72$, $p < 0\ .001$). Vocabulary and education also strongly correlated ($r = 0.72$, $p < 0.001$).

To determine how strongly each of these variables is related to performance on the nonce-word inflection task, partial correlations were computed first for vocabulary and inflection with education held constant and then for education and inflection with vocabulary held constant. The partial correlation between vocabulary and inflection with education held constant was no longer significant ($r = 0.28$, $p = 0.107$), while the correlation between education and inflection with vocabulary held constant remained significant ($r = 0.48$, $p = 0.003$). Thus, the differences in performance on the inflection task are more directly dependent on differences in education than on differences in vocabulary size.

Why should performance on the inflection task depend on education? We can rule out relatively uninteresting reasons such as lack of familiarity with the testing situation, willingness to cooperate, or failure to understand the experimental task, since all participants reliably supplied the dative forms of some nouns, e.g. feminine nouns from densely populated neighbourhoods. It is also extremely unlikely that the educated

participants were relying on explicit knowledge acquired during grammar lessons at school, since none of a smaller group of educated adults who were questioned about this were able to formulate the rules for forming the dative. Since the experimental task depends to a certain degree on metalinguistic skills, which are likely to be better developed in the more educated participants, one could argue that such skills are at least partially responsible for the observed differences. However, the fact that all virtually all participants performed extremely well on masculine and feminine nouns from high density neighbourhoods (masculines: mean 89%, median 100%; feminines: mean 95%, median 100%) suggests that this is not the case. Thus, we must conclude that the differences observed in the experiment reflect genuine differences in linguistic knowledge.

Follow-up studies designed to pinpoint the source of the less educated participants' difficulties showed that they reliably inflected real nouns after *dzięki* and *przyglądać się* (so they knew which case was required in the dative contexts used in the experiment) and had no problems selecting the gender-appropriate form of a demonstrative adjective used in construction with the nonce nouns (showing that their failure to provide the correct dative inflection could not be due to problems with identifying the gender of the nonce noun). Thus their relatively low scores on the inflection task appear to be attributable to problems with the inflections themselves.

The effect of education can be most plausibly attributed to asymmetries in the distribution of dative nouns in spoken and written discourse, and differences in the amount of exposure to written discourse. In spoken texts, the dative case is used to mark semantic functions such as recipient, beneficiary, addressee, and experiencer. All of these are typically animate roles, and hence the vast majority of dative nouns in spoken texts are

animate nouns, typically kinship terms and personal names, resulting in relatively low type frequencies for individual endings. The dative can also be used with inanimate nouns after certain prepositions (*ku* 'towards', *wbrew* 'notwithstanding', *przeciw(ko)* 'against'), verbs (e.g. *sprzeciwiać się* 'oppose', *ulegać* 'succumb, submit', *dorównywać* 'to equal') and adjectives (*wierny* 'faithful', *wrogi* 'hostile', *bliski* 'close'). All of these uses, however, are fairly literary, even archaic, and hence largely restricted to written texts. The effect of this is that the dative occurs in a much wider range of constructions, and hence with a much wider variety of nouns, in written language. One simple measure of this difference in distribution is the number of inanimate nouns used in the dative as a proportion of all dative-marked nouns in various genres. This ranges form 1.4% in child-directed speech to 14% in adult-directed speech and 62% in written texts.[3] Since more educated speakers have more experience with formal written language, they encounter a larger number of noun types in the dative; and since exposure to a larger number of different noun types with a particular inflection results in greater productivity (see above), more educated speakers are more likely to use the dative inflections productively.

It should be stressed that the individual differences observed in this study are not due to linguistically irrelevant performance factors such as failure to understand the task or

[3] These figures are based on the Marysia corpus (which consists of transcripts of a thirty-hour sample of the linguistic experience of a two-year-old Polish girl collected by the author), the Otwinowska-Kasztelanic (2000) corpus of spoken Polish and a random sample of 200 nouns from the IPI-PAN corpus of written Polish (available from http://korpus.pl/en/), respectively.

uncooperativeness, since all participants performed at or close to ceiling on some subcategories of words. They are also not attributable to dialectal differences, as virtually all varieties of Polish form the dative in the same way, and all participants reliably supplied the target endings with real words. They reflect genuine differences in productivity of the dative inflection. Moreover, productivity is clearly a matter of degree. All but one of the participants were productive with all of the endings, in the sense that they were able to correctly inflect at least one nonce word of each gender. However, many were not consistently productive: they supplied the correct ending with some nonce nouns belonging to a particular gender, but not with others.

Four complex English constructions

Both of the examples discussed so far involved knowledge of morphological constructions. Can similar differences be observed for syntactic knowledge – which, according to many linguists, depends critically on a shared innate universal grammar? Dąbrowska (1997) tested comprehension of four types of sentences (all relexified examples from *Linguistic Inquiry*): complex NP sentences, which contained a subordinate clause with a noun complement clause in the subject position (2), 'tough movement' sentences (3), and two types of sentences with parasitic gaps (4-5). Interspersed with the test items were control sentences which were slightly longer, but did not contain any difficult structures (6).

(2) Paul noticed that the fact that the room was tidy surprised Shona.

(3) John will be hard to get his wife to vouch for __.

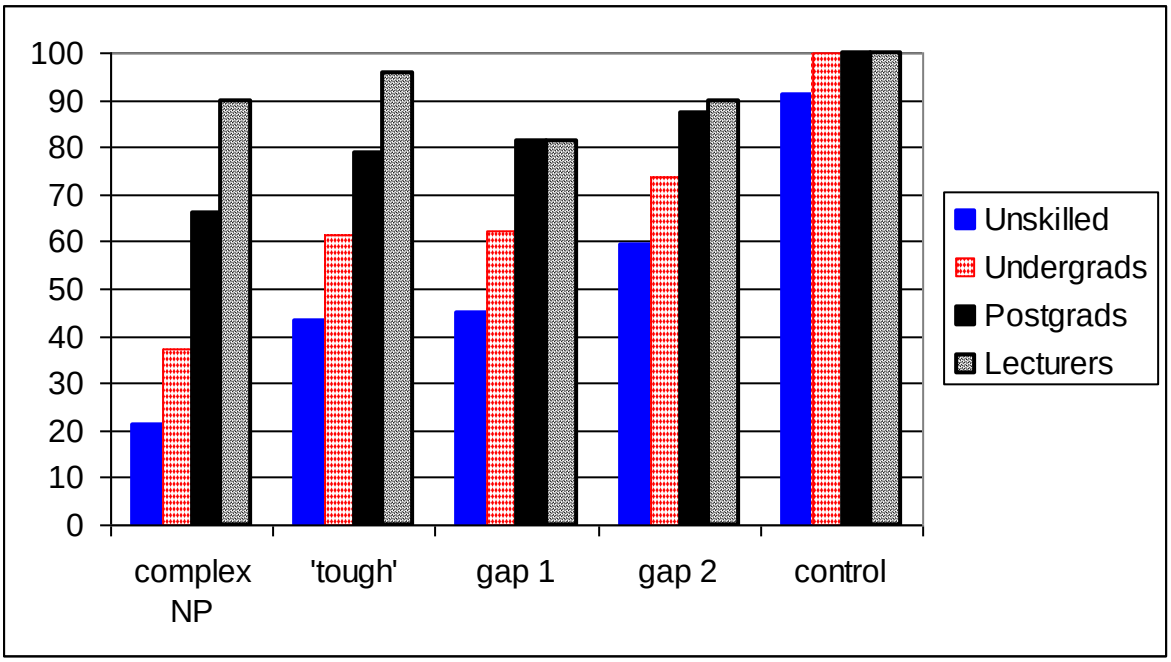
(4) The nervous-looking student that Chris met __ after being told his girlfriend wanted to jilt __ took the 11 o'clock train.

(5) It was King Louis who the general convinced __ that this slave might speak to __.

(6) Alex decided that the easiest way to find out whether or not the plan would work was to ask the man who played the guitar at the party.

The participants were cleaners, janitors, undergraduate and postgraduate students and lecturers from the same university. There were 10 participants in each group; since the results for the cleaners and janitors were virtually identical, the data for these two groups were pooled in the analyses described below. In order to ease the demands on working memory without disadvantaging the less educated participants (who may have had poor reading skills), the test sentences were presented orally and in writing; the participants could ask to have the sentences repeated if they wished (and the less educated participants frequently did). The participants listened to the sentences and then answered simple questions about them. For instance, the questions for the complex NP sentence in (2) were *What did Paul notice?* and *What surprised Shona?*. *That the fact that the room was tidy surprised Shona*, *that Shona was surprised*, or some paraphrase of this counted as a correct answer for the first question; an example of an incorrect answer would be *that the room was tidy*. For the second question, the correct answer was *that the room was tidy* or *the fact that the room was tidy*; an example of an incorrect answer would be *that Paul noticed*.

Figure 4: Education-related differences in comprehension of four complex construction

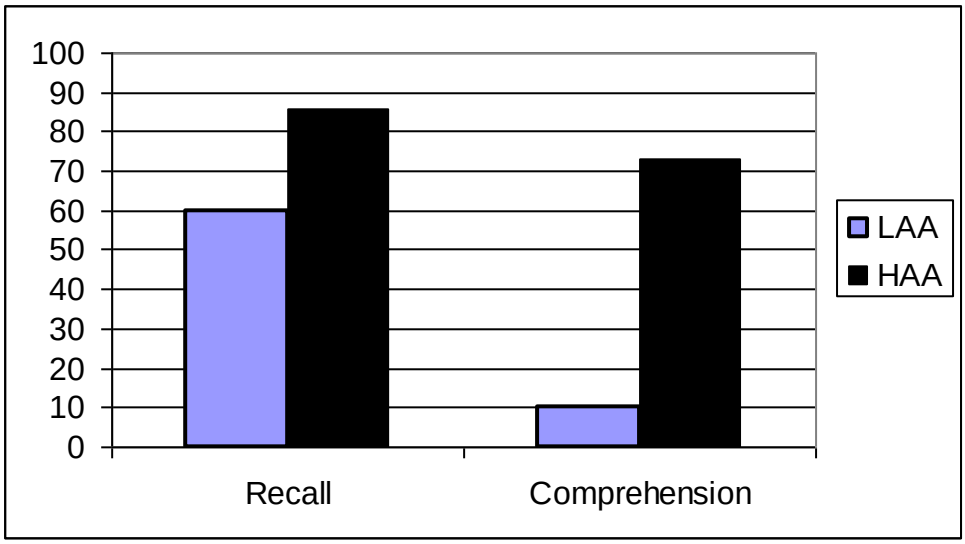


As can be seen from Figure 4, the experiment revealed very robust education-related differences for all four experimental conditions, but not for the control sentences. Are these differences attributable to competence or performance? It should be noted that the participants were tested under ideal conditions: the sentences were presented to them in both spoken and written form, and they were given as much time as they needed – so there is a real sense in which some participants' inability to give the correct response can be seen as a problem with linguistic knowledge, i.e. competence, rather than the ability to access that knowledge. On the other hand, there is no denying that the structures used in the experiment placed heavy demands on the processing system, and hence one could cogently argue for a performance explanation of the differences.

This issue was addressed by Chipere (2001), who tested comprehension and recall of complex NP sentences by two groups of speakers: Low Academic Achievement (LAA) and High Academic Achievement (HAA) participants. The LAA group consisted of eighteen 18-year-olds who got a D or below in GCSE English; the HAA group consisted of eleven 18-year-olds from the same school who got A's in at least 5 GCSE subjects, including English.[4] Comprehension was tested by means of simple questions similar to those used by Dąbrowska (1997). In the recall task, participants were presented with sentences on a computer screen one word at a time and asked to repeat the whole sentence after the last word. As can be seen from Figure 5, the HAA group performed considerably better on both tasks.

[4] GCSE (General Certificate of Secondary Education) is an academic qualification in England, Wales and Northern Ireland. GCSEs are normally taken at age 16 in a number of subjects. There are 8 pass grades (A*, A, B, C, D, E, F, G); grades A*-C are considered good passes. Typically, receiving five or more A*-C grades is a prerequisite for proceeding to the next level, i.e., General Certificate of Education Advanced Level (commonly referred to as "A-level").

Figure 5: Pre-test results (redrawn from figures given in Chipere 2001)



These results are compatible with both explanations mentioned above. One could argue that the LAA participants had not mastered the complex NP construction and hence found it difficult to remember the experimental sentences, since it is difficult to remember unstructured material. Alternatively, one could argue that they were unable to hold the test sentences in working memory, and hence could not understand them.

In the second part of the experiment, the LAA participants were randomly assigned to one of two groups: a memory training group, who learned to repeat complex NP sentences, and a comprehension training group, who were given a brief explanation of this structure followed by practice with feedback. Both groups were trained on the pre-test sentences and then tested with a new set of sentences. The results of the post-test show that memory training improves performance on the memory task, but has no effect on the comprehension task (see Figure 6). Comprehension training, on the other hand, improves comprehension: participants in this group are now at ceiling, better than the HAA group on

the pre-test (see Figure 7). It also improves memory: the comprehension-training group's performance on the recall post-test was similar to that of the HAA group on the pre-test. Chipere concludes that the LAA speakers' poor performance on the pre-test is attributable to lack of linguistic knowledge, not limited processing capacities. (For further discussion of this issue, see the section on competence and performance below.)

Figure 6: Post-test results for the memory training group

(redrawn from figures given in Chipere 2001)

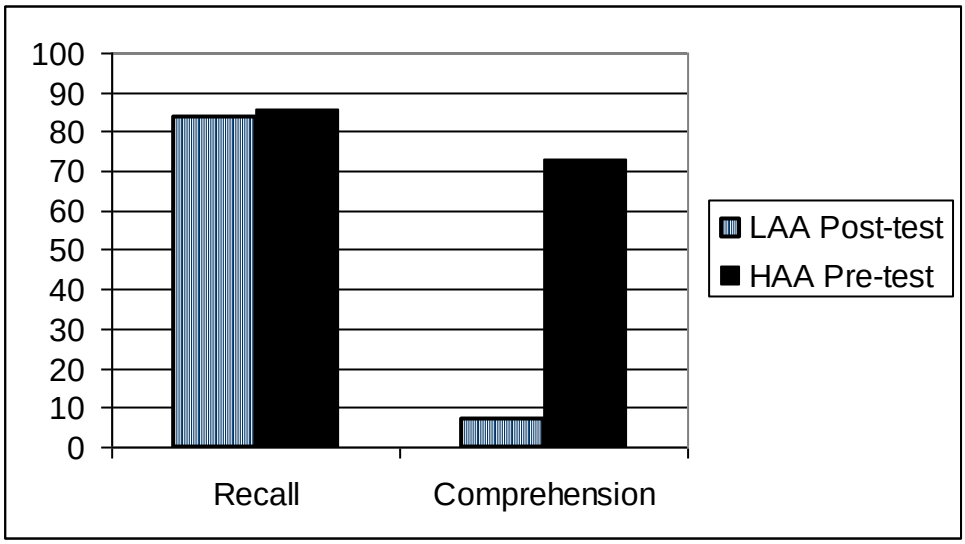
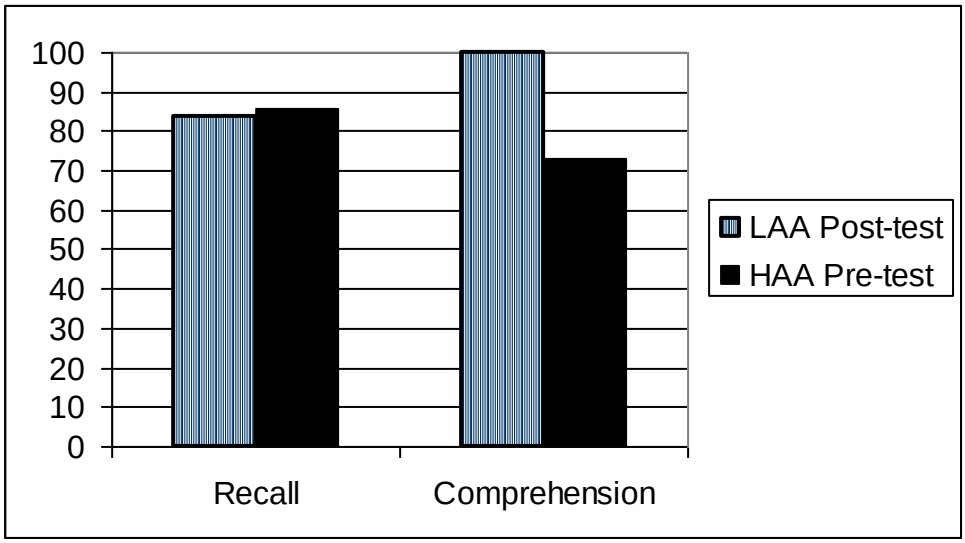


Figure 7: Post-test results for the comprehension training group

(redrawn from figures given in Chipere 2001)



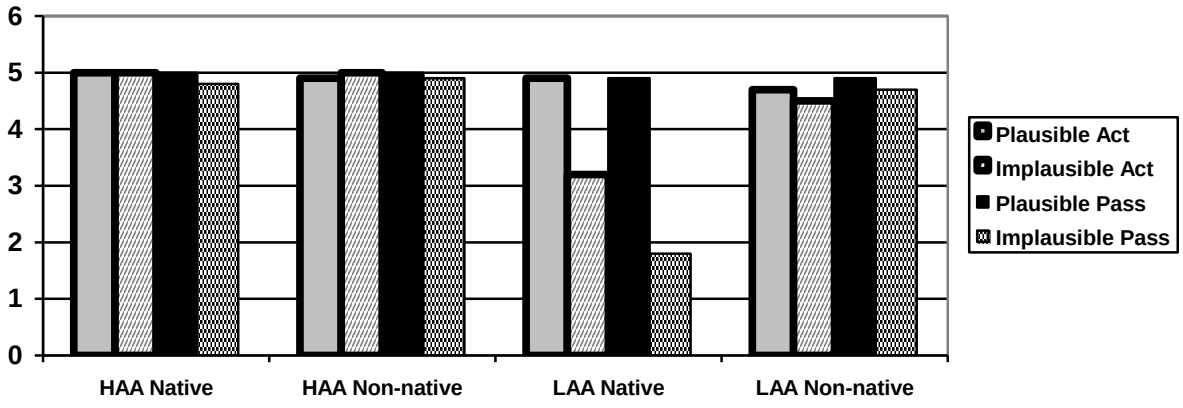
Passives

The studies described in the preceding section tested comprehension of very complex structures which place heavy demands on working memory. Are there similar differences in the ability to comprehend simpler sentences? Dąbrowska and Street (2006) tested comprehension of four sentence types: plausible passives (*The man was bitten by the dog*), implausible passives (*The dog was bitten by the man*), plausible actives (*The dog bit the man*), and implausible actives (*The man bit the dog*). Participants were asked to listen carefully to the test sentences and identify the 'do-er' (i.e. agent). The experiment was a partial replication and extension of an earlier study by Ferreira (2003), who tested psychology undergraduates and found that they were at ceiling on actives but had problems with passives, particularly implausible passives, where they chose the correct noun 74% of

the time – well above chance, but nevertheless far from perfect. Ferreira explains this by suggesting that instead of doing a full parse, speakers sometimes rely on what she calls ‘good enough’ representations (see below for further discussion).

The participants tested by Ferreira were undergraduate students. Since full passives are found mostly in formal written texts, undergraduates would have encountered quite a large number of instances of this construction. If proficiency with a particular construction depends on the amount of experience with that construction, individuals without much formal schooling might have even more problems with the passive, while more literate individuals might do better. To investigate this possibility, Dąbrowska and Street compared the performance of a highly educated group (British post-graduate students, all of whom had at least 15 years of formal education) with that of less-educated individuals (checkout assistants and shelf-stackers at a supermarket who had no more than secondary-school education). In addition, in order to determine whether the type of experience with a particular structure – as opposed to sheer quantity – also matters, they tested two additional groups: highly-educated and less-educated non-native speakers. There were 10 participants in each group.

Figure 8: Comprehension of active and passive sentences (Dąbrowska and Street 2006)



As can be seen in Figure 8, both high academic achievement groups were at ceiling in all four conditions. The native non-graduates, on the other hand, had problems with implausible sentences, particularly passives, where they were at chance. Most surprisingly, non-native non-graduates were also at ceiling on all sentence types: that is to say, they performed better than the less-educated native speakers.

How can we explain the less-educated native speakers' poor performance? One obvious possibility is that they were unfamiliar or uncomfortable with the testing situation, or simply uncooperative. However, their good performance on plausible sentences argues against such an interpretation, as does the fact that the less educated non-natives – who were presumably even more uncomfortable, since they were tested in a foreign language in a foreign country – did extremely well. It is also worth noting that the participants in Ferreira's study, who were psychology students and thus presumably no novices to testing, also had problems with implausible passives.

Another possibility is that they had misunderstood the task, and thought that they were being asked what someone who produced the test sentences probably meant. Thus, on being confronted with a sentence describing an extremely improbable event, they assumed that the speaker had made a mistake, and gave an answer that was scored as incorrect. This kind of pragmatic normalisation may well have affected performance on the experimental task; however, it cannot account for the full pattern of results observed in the study. In particular, it does not explain the difference in performance on implausible actives and passives, or why it was only the less educated natives who normalised. Furthermore, pragmatic normalisation cannot explain Ferreira's findings: in her experiment, performance on neutral passives (79%) was only slightly better than on implausible passives (74%), while performance on implausible actives was at ceiling (99%).

A more promising explanation would be one appealing to processing capacity. We know that passive sentences are difficult to process because of their non-canonical word order. Implausible sentences are also difficult to process because the formal and pragmatic cues are in conflict. Implausible passives thus place particularly heavy demands on the processing system, and may be too difficult for speakers with limited processing resources. Since educational achievement is known to correlate with working memory capacity (Gathercole et al. 2004), differences in working memory could in principle be responsible for the education-related differences observed in the comprehension task.

However, such an account is problematic for several reasons. As pointed out earlier, Chipere's LAA group performed at ceiling on a more complex structure after training, which suggests that their original problems with these structures – and, presumably, also the non-graduate group's problems in this experiment – were not due to limited memory

capacity. Secondly, the undergraduate students in Ferreira's study also performed poorly on passives, and one would hardly want to attribute this to processing limitations, since it would imply that only a very small proportion of the population are able to process passive sentences. Finally, as we will see shortly, two further studies found education-related differences in comprehension of passive sentences using materials that did not involve a conflict between syntactic and pragmatic information.

Dąbrowska and Street conclude that the less educated speakers' difficulties with passives are attributable to the fact that they have had less experience with them (since passive sentences are relatively infrequent in spoken language) and hence lack a well-entrenched passive schema. However, the sheer amount of exposure cannot be the only reason for the observed differences, since it cannot explain why the less educated non-native speakers performed better than the less-educated natives, who presumably had more exposure to passives. Clearly other factors are at work. These could include the type of experience (explicit instruction, exposure to a relatively large number of exemplars over a relatively short period), metalinguistic skills (presumably enhanced in the non-native group by explicit instruction in second language), or differences in motivation and ability.

Further evidence for education-related differences in native speakers' mastery of the passive construction comes from a more recent study by Street and Dąbrowska (submitted) which tested comprehension of semantically reversible sentences using an online task. There were two groups of participants. The HAA group (N=31) were postgraduate students or recent graduates from a variety of academic backgrounds; all had at least 17 years of formal education. The LAA participants (N=32) were employed as packers, cleaners or hairdressers and had at most 11 years of formal education. Participants were presented with

simple sentences such as *Sally was bitten by Rachel* on a computer screen. Each sentence was followed by one of the NPs mentioned in the sentence (i.e., *Sally* or *Rachel*); the participants' task was to decide whether this person was the 'do-er' (i.e., the agent) or the 'acted-on' (i.e., the patient). The test consisted of 12 actives and 12 passives with the same verbs. Half of the verbs contained active-attracting verbs (verbs which are used almost exclusively in the active voice) and the other half passive-attracting verbs (verbs which occur relatively frequently in the passive voice), as determined by collostructional analysis (Gries and Stefanowitch 2004).[5] The purpose of this manipulation was to establish whether the education-related differences in performance observed in the earlier study could be attributed to the LAA group relying on lexically specific representations (e.g. *NP1 BE bitten by NP2*) rather than an abstract passive construction (*NP1 BE VERB-ed by NP2*).

There were two dependent variables: decision accuracy and reaction time. The decision accuracy results were similar to those observed in the earlier study. Both groups were at ceiling on actives (HAA: 99% correct; LAA: 98% correct). The HAA participants were also virtually at ceiling on passives (96% correct). The LAA participants, on the other hand, chose the correct response on only 86% of the passive stimuli. Thus, *as a group*, their performance on passives was above chance but significantly worse than that of the HAA group, or their own performance on active sentences. However, within the group, there were considerable individual differences. 10 participants (i.e., 31%) were at ceiling,

[5] Since active sentences are much more frequent then passives, passive-attracting verbs are not necessarily more frequent in the passive than the active; they are, however, attested more frequently in this construction than expected, given the frequency of the passive.

supplying the target response on all 12 of the passive test items; and another 14 (44%) performed above chance but below ceiling. Of the remaining participants, 7 were at chance, achieving scores between 3 and 10 out of 12, and 1 was below chance, supplying only 1 correct response on the passive trials. (This participant scored 11/12 on active sentences, so s/he consistently interpreted the first NP in the sentence as the agent.)

The reaction time results revealed a somewhat different pattern. The HAA participants were faster than the LAA group on all sentence types. This is not very surprising, since the test sentences in this study were presented in writing; thus, the faster reaction time may simply be indicative of the fact that the more educated participants read faster. It is also possible that they are able to process all kinds of linguistic information faster, whether it is presented orally or in writing. Furthermore, both groups were faster on actives than on passives, and both groups were faster on passives with passive-attracting verbs than on passives with active-attracting verbs, which suggests that both HAA and LAA speakers have lexically specific schemas for verbs which are frequently used in the passive. Crucially, there was no interaction between group, construction and verb, or between group and construction on either decision accuracy or reaction time, which indicates that the less educated participants' relatively poor comprehension of passive sentences cannot be explained by assuming that they rely primarily or exclusively on low-level schemas.

Quantifiers, and passives again

The final example of individual differences in grammatical knowledge that will be discussed here is knowledge about quantifier scope, in particular, the interpretation of sentences with the universal quantifier *every.* Two earlier studies by Brooks and Sekerina (2005/2006, 2006) suggest that such knowledge is acquired late in the course of acquisition (after age 9), and that even adults sometimes misinterpret sentences such as (6) and (7). In the first study, undergraduate students were presented with an array of pictures and asked to choose the one that went with the sentence; they chose the correct picture 79% of the time. In the second study, another group of undergraduates listened to similar sentences and had to decide whether they matched a particular picture; the average number of correct responses was 87%. Thus, in both studies, group performance was above chance but not at ceiling; and within each group, some individuals were in fact performing at chance.

(6) Every fish is in a bowl.

(7) Every bowl has a fish in it.

Street and Dąbrowska (2010) describe two experiments designed to determine whether such comprehension problems are related to educational achievement and whether additional experience with the construction leads to improved performance. The experiments tested comprehension of the two variants of the quantifier construction exemplified above: the locative variant (Q-*is*) and the 'possessive locative' (Q-*has*), both with the quantifier *every* modifying the subject noun. In addition, the test also included

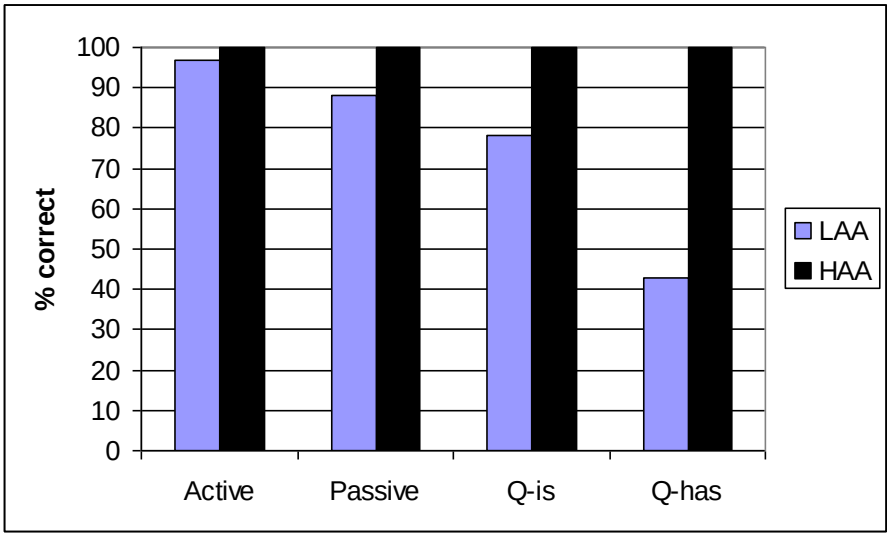
unbiased passives such as (8), included to determine whether the education-related differences observed by Dąbrowska and Street (2006) could be replicated using a different testing method and materials that did not contain implausible sentences, and the corresponding actives (9), which were a control condition. There were thus four conditions: Q-*is*, Q-*has*, Passive, and Active.

(8) The girl was kissed by the boy.

(9) The boy kissed the girl.

The first experiment tested two groups of native English speakers. The high academic achievement group (N=19) were postgraduate students at a university in the north of England; all had at least 17 years of formal education. The low academic achievement group (N=31) had at most 11 years of formal education and were employed as shelf stackers, packers, assemblers or clerical workers. The participants were asked to listen carefully to the test sentences and then choose one of two pictures which went with the sentence. For the active and passive sentences, the pictures depicted simple reversible transitive events (e.g. a boy kissing a girl and a girl kissing a boy). For the quantifier sentences, they depicted objects and containers in a partial one-to-one correspondence (for instance, three bowls containing a fish plus one empty bowl, and three bowls containing a fish plus one fish without a bowl). The results are presented graphically in Figure 9.

Figure 9: Experiment 1 results (Street and Dąbrowska 2010)



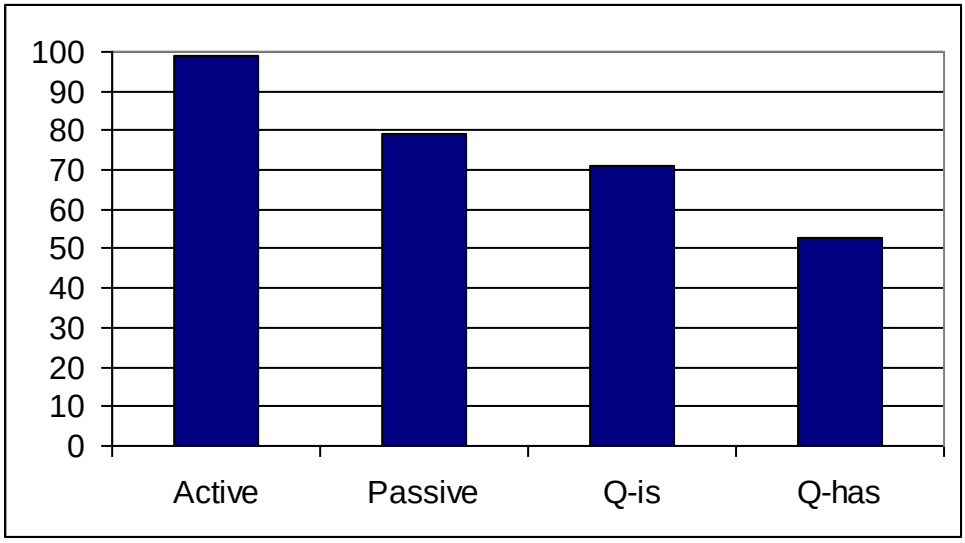
As can be seen from the figure, the HAA participants performed at ceiling: in fact, all graduate participants scored 100% in all conditions. The LAA group were also at ceiling on actives (97%), somewhat worse on passives (88%), worse still on Q-*is* sentences (78%), and at chance on Q-*has* (43% correct). This order of difficulty reflects the relative frequency of the four constructions: the British National Corpus, a 100 million corpus of contemporary British English, contains over 120,000 actives, 5675 full passives, 8 instances of Q-*is* sentences, and no instances of Q-*has*. Thus, the differences in performance on these four constructions can be plausibly explained in terms of entrenchment: speakers develop stronger representations for constructions they have more experience with.

Experiment 2 was a training study. The participants were 54 adults enrolled in a Skills for Life course at a local college. Skills For Life courses are aimed at adults who left

school without any formal educational qualifications. They comprise five levels: three entry levels plus Levels 1 and 2. Entry level courses are intended to teach basic literacy skills such as reading a newspaper article or instruction manual and writing a letter to a utility company. Level 1 is equivalent to a GCSE pass, and Level 2 to a good pass (C or above) at GCSE (see footnote 4 for information about GCSEs).

The experiment consisted of six stages: pre-test, training, three post-tests, and follow-up questionnaires. The tests were sentence-picture matching tasks similar to those used in experiment 1. There were four versions of the test, each containing sentences with the same verbs but different noun phrases; the order of presentation of the four versions was counterbalanced across participants and stages.

Figure 10: Pre-test results (Experiment 2, all participants)



The results for the pre-test were very similar to those of the non-graduate group from the first study: the participants obtained very high scores on actives, had some

problems with passives, still more problems with Q-*is* sentences, and were at chance on Q-*has* (see Figure 10). The pre-test results were used to select low-scoring participants, i.e. individuals who obtained scores of no more than 4 out of 6 in all three experimental conditions. Seventeen participants met this criterion and were randomly assigned to either the passive training or the quantifier training group. The training involved an informal explanation of the target structure followed by a practice session during which participants did the pre-test again, and were given feedback on their performance after every item. The practice session was immediately followed by post-test 1. Post-tests 2 and 3 were administered one week and twelve weeks after training respectively. No feedback was provided during the post-tests. Finally, all participants, including those who had not been selected for the training phase, were given a questionnaire investigating their reading habits and the short version of the Need for Cognition questionnaire (Cacioppo, Petty and Kao 1984), which measures how much people enjoy effortful cognitive activities.

The results for the passive training group (N=8) are presented in Figure 11, and show a clear improvement in performance on the passive, but not the quantifier constructions, after training. The results of post-tests 2 and 3 show that the effects of training are long-lasting. The quantifier-training group (N=9), in contrast, shows a clear improvement in performance on sentences with quantifiers, and no change in performance on passives (see Figure 12). Again, the effects were long-lasting.

Figure 11: Passive training group results (Street and Dąbrowska 2010)

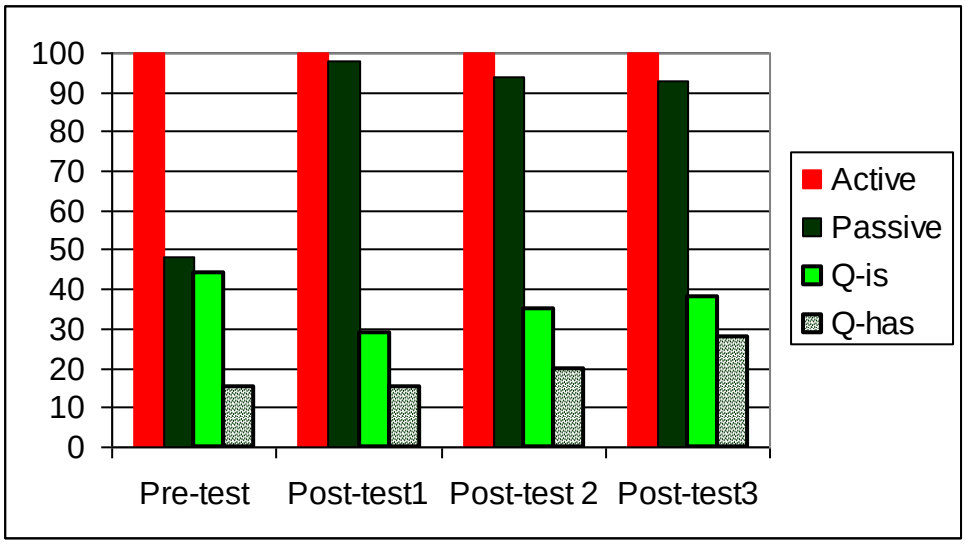
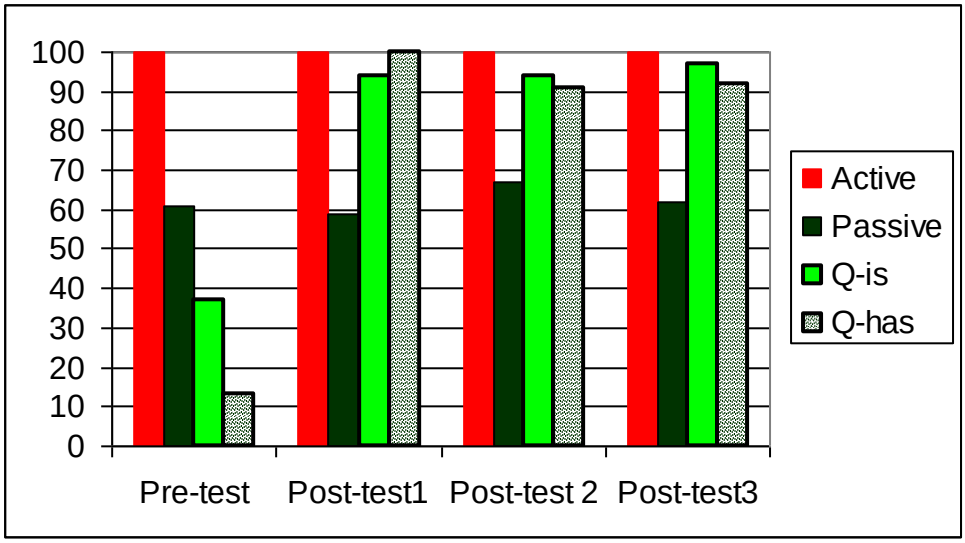


Figure 12: Quantifier training group results (Street and Dąbrowska 2010)



Thus, performance improved dramatically as a result of additional experience with the relevant construction, showing that the initial differences in test scores are attributable to differences in familiarity with specific linguistic constructions. These results raise an

interesting question. If the LAA participants are able to learn the constructions used in the experiments after a fairly minimal amount of training (6 exemplars), why hadn't they learned it earlier? Although full passives and the variants of the quantifier construction used in the experiment are relatively infrequent (see above), all the participants are likely to have experienced considerably more than 6 exemplars of each in their lifetimes, and hence, one could argue, should have acquired the relevant knowledge before the experiment began.

The training provided during the experiment differs from normal exposure in two ways: it was more intensive (the participants were presented with a number of exemplars in a very short time) and it involved explicit explanation and feedback as well as exposure to relevant data. Research on construction learning (and learning in general) shows that 'spaced' exposure, where individual learning episodes are distributed over a number of sessions, is more effective than 'massed' exposure, where the same number of learning episodes occurs in a single session (Ambridge et al. 2006). This suggests that the fact that the training session provided more intensive exposure is unlikely to be responsible for the difference in results.

Thus, it seems that the critical factor was the *type* of input available to the learner. Not every instance of exposure to the relevant structure is necessarily a learning episode. In order for construction learning to take place, there must be enough contextual information to allow the learner to infer the meaning of the sentence, and the learner needs to be attending to the linguistic form as well as the contextual information. In the training phase of the study, the experimenter explicitly directed the participants' attention to the relevant aspects of both form and meaning, thus maximising the chances of learning taking place. It is conceivable that such favourable circumstances did not occur often enough in the

participants' prior experience, either because there was not enough contextual information or because they were not attending to it.

Competence or performance?

The studies summarised here suggest that there are considerable differences in adult native speakers' ability to produce and comprehend a variety of linguistic structures. This raises an important question, namely, whether the observed differences between speakers – and in particular, the education related differences – reflect differences in underlying knowledge (i.e., linguistic competence), or whether they are better thought of as differences in linguistic performance.

Is it possible, for instance, that the less educated participants' relatively low scores could be explained by the non-linguistic task demands, or simply failure to cooperate with the experimenter? We know from studies of language acquisition that children's performance in experimental settings often varies as a function of task: children who fail on experiments testing knowledge of a particular structure sometimes do very well when tested using a different method. Could this also be true of the LAA participants – in other words, would they have performed better if they were tested in a different setting or using a different method, or if they were more motivated?

We can be confident that the less educated participants had understood the experimental tasks and engaged with them. This is evident from the fact that they performed very well in the control conditions, and also from other aspects of their behaviour during the experiments: they frequently asked the experimenter to repeat the test

sentences or repeated the test sentences to themselves several times before responding; they thought carefully before responding and occasionally argued with the experimenter, suggesting either that the test sentences were ambiguous (in the Dąbrowska 1997 study), or that neither of the two answers given was possible (in the picture selection task used in Street and Dąbrowska 2010). There is also every reason to believe that they were able to perform the experimental task, since, again, they performed at or near ceiling in the control conditions. Furthermore, it is important to note that – with the exception of the doer-or-acted-on task used by Street and Dąbrowska (submitted) – the tasks used in these studies are not very demanding cognitively. Picture selection and nonce word inflection tasks are used with children as young as two, and comprehension questions with children who are only slightly older – and the participants in these studies were adults with no known learning difficulties. Finally, when the same structure – i.e., the passive – was tested using different methods, this did not result in substantial differences in performance: the proportion of target responses in the picture selection task used in Street and Dąbrowska (2010) – 88% in Experiment 1 and 79% in Experiment 2 (pre-test) – is very similar to that observed in the much more difficult task used by Street and Dąbrowska (submitted) – 85%.[6] (Note, too, that the test sentences in these two studies were presented in different

[6] The proportion of target responses for passive sentences observed in Dąbrowska and Street (2006) was different – but this, as indicated earlier, is due to the fact that participants tended to choose the pragmatically most plausible response, which resulted in high scores on the plausible sentences and very low scores on implausible sentences.

modalities: orally in the former study and in writing in the latter one; again this did not make any difference.)

An alternative possibility is that the participants had understood the task and had the linguistic knowledge necessary to perform it, but were unable to access this knowledge during the experiment. Such an explanation, of course, is vacuous unless we have a clear account of what it was exactly that prevented participants from accessing their knowledge. As pointed out earlier, one reasonable possibility is limited working memory capacity. During sentence processing, an individual must simultaneously maintain the incoming linguistic material in memory, retrieve information from long-term memory, and perform various computational operations on the two. It is well-known that there are correlations between comprehension of some complex grammatical structure and working memory capacity. Accordingly, some researchers (see e.g. Just and Carpenter 1992, King and Just 1991, Waters and Caplan 1996) have proposed that working memory directly constrains syntactic comprehension, so that individuals with low working memory capacity (as measured by the reading span task) may have difficulty processing more complex structures, particularly if they also have to perform a concurrent task. Since working memory capacity correlates with educational achievement (Gathercole et al. 2004), we have an alternative explanation of the less educated participants' relatively poor performance in the experiments described here.

There are two reasons for being sceptical about the feasibility of such an account. First, while differences in working memory may offer a plausible explanation of the results of the comprehension experiment reported in Dąbrowska (1997), which tested relatively complex sentences, it is doubtful whether it could be applied to the more recent studies

(Dąbrowska and Street 2006, Street and Dąbrowska 2010, submitted), which tested comprehension of 6-7 word sentences with no embedding such as *The boy was kissed by the girl* and *Every bird is in a cage;* and it certainly cannot account for the differences in performance on the nonce word inflection task described in Dąbrowska (2008b), where, as explained earlier, participants who failed to correctly inflect nonce nouns from low-density neighbourhoods were able to apply the same ending with other types of nonce nouns. Furthermore, a working memory account is difficult to reconcile with the dramatic improvement in performance observed in the two training studies (Chipere 2001 and Street and Dąbrowska 2010, Experiment 2) if we adopt the traditional view of working memory as a resource with a fixed capacity. It seems that a more promising explanation for the correlation between performance on the reading span task and syntactic comprehension is that proposed by MacDonald and Christiansen (2002), who argue that it is attributable to the fact that the reading span task actually measures linguistic proficiency rather than working memory capacity. Chipere's (2001) finding that comprehension training improves performance on the sentence recall task also supports this view, as do the results of a training study by Wells et al. (2009) in which low-span individuals, after intensive exposure to relative clauses, developed a reaction time profile characteristic of high-span individuals.

The final possibility is that the LAA participants have the same linguistic representations as the HAA group, but use different parsing strategies – specifically, they only carry out a superficial analysis of the sentence. According to some processing models (e.g. Late Assignment of Syntax Theory – see Townsend and Bever 2001) sentence processing involves two distinct phases. During the first phase, the processing system uses

various probabilistic cues and heuristics to construct a “quick and dirty” pseudo-parse, which is then used to guide the true parse – a slower algorithmic process which accesses syntactic knowledge to construct a complete syntactic representation. Constructing the true parse is computationally quite demanding, so this stage of analysis may not always be carried out, especially under time pressure, or when processing resources are limited; in such cases, listeners will rely on the pseudo-parse to arrive at the sentence’s meaning. Indeed, Ferreira (2003; see also Ferreira, Bailey and Ferraro 2002 and Ferreira and Patson 2007) has argued that this happens quite often: the representations constructed by the pseudo-parse are “good enough” for most everyday purposes. For instance, applying the NVN strategy (Noun-Verb-Noun = Agent-Verb-Patient – see Bever 1970, Townsend and Bever 2001) will enable the comprehender to assign the correct interpretation to most English sentences. It results in an incorrect interpretation for sentences with non-canonical word order (such as passives), but in real life, most passives are semantically non-reversible, so comprehenders can arrive at the intended interpretation by relying on pragmatic cues. It is not implausible to assume that less educated speakers are more likely to rely on the pseudo-parse, and this could explain at least some of the observed group differences – specifically, the differences in performance on passives (Dąbrowska and Street 2006; Street and Dąbrowska 2010) and possibly also sentences with quantifiers and the complex constructions studied by Dąbrowska (1997).

Note that, since constructing the true parse involves an additional processing stage and since this second stage is, by hypothesis, considerably slower than the pseudo-parse, the two-stage processing account makes a clear prediction: participants who construct the true parse, and hence interpret the sentences correctly, should be slower than participants

who only use processing heuristics – in other words, there should be a speed-accuracy trade-off. This prediction was tested in Street and Dąbrowska (submitted), who found the exact opposite: the more accurate participants were also faster, both within groups and in the sample as a whole, and, within participants, there was no significant difference in reaction times for correct and incorrect trials. Thus, the differences in performance on the passive cannot be explained by appealing to processing strategies.

Reasons for individual differences

The studies described here demonstrate that there are considerable individual differences in performance on tasks tapping knowledge of basic linguistic constructions, including case marking, 'tough movement', complementation, passives, and quantifiers. These differences cannot be explained by appealing to working memory capacity, test-taking skills, or willingness to cooperate with the experimenter. They are, however, strongly correlated with education.

Why should there be a relationship between education and differences in grammatical knowledge? One possibility that has already been hinted at is that more educated speakers have more experience with written language. Partial support for this suggestion can be gleaned from the Street and Dąbrowska (2010) study, which found a significant correlation between amount of reading and performance on the initial comprehension test among adult literacy students (*rho* = 0.551, $p < 0.001$). A greater amount of experience with written language could also explain some of the group differences, specifically, the educated participants' better performance on passives and

complementation structures (which are more frequent in written language) and possibly also the dative case (which, as argued earlier, occurs in a wider variety of constructions, and hence with more noun types, in written texts). It does not, however, explain the differences in performance on sentences with the universal quantifier *every* which is, if anything, more frequent in spoken language than in written texts (see Street and Dąbrowska 2010).

It is also possible that the postgraduate students had more linguistic experience *tout court* than the less educated participants. Graduate participants rely on language more in their daily lives than individuals who do menial jobs, in that most of their working day is spent in some kind of linguistic activity. They also tend to read more, and are more likely to be skilled readers, and since skilled readers absorb more words per unit of time than skilled conversationalists (see Dąbrowska 2004: 19), they are likely to have had more exposure to all the structures tested than the non-graduates. Finally, a high proportion of postgraduate students come from professional families, and since professional parents tend to talk more to their children than working-class parents (cf. Hart and Risley 1995, 1999), they are likely to have experienced more language as children.

A third possibility is that the participants' prior linguistic experience was qualitatively different. As suggested earlier, the improvement in performance observed in Street and Dąbrowska's (2010) second experiment and Chipere's (2001) study is probably at least partly attributable to the fact that participants were made to attend to both form and meaning at the same time during the training phase. Interestingly, in both studies a number of participants reported having an "aha" experience when the construction was explained to them: they admitted they had been guessing during the pre-test but relied on their newly

acquired knowledge on the post-tests, and their performance on the post-tests (and the practice test used during training) corroborates this. It is conceivable that the more educated participants had had parents (or teachers) who were more likely to provide this kind of explicit instruction in language, and hence had such an "aha" experience earlier in life.

It is widely acknowledged that explicit learning plays a significant role in L2 learning, at least at the level of "noticing" (Schmidt 1990); however, the general consensus is that grammatical development in L1 is virtually entirely implicit, largely because young children are believed to lack the metalinguistic skills necessary to benefit from explicit explanations of grammar. The results of the two training studies suggest that we may need to revise our views on the role of explicit learning and teaching in first language development. L1 and L2 learning may not be as different as generally believed in this respect – although the mix of explicit and implicit mechanisms presumably is different (see Dąbrowska 2009).

All of the explanations discussed so far attribute education-related differences in language attainment to characteristics of participants' prior language experience. However, it is also possible that the education-related differences in linguistic proficiency are attributable to characteristics of the learner rather than the external environment: that is to say, it is possible that the more educated participants are more able or more motivated to learn and hence more successful at language learning as well as learning in an academic context. There is some evidence supporting both of these hypotheses. Thus, Brooks and Sekerina (2006) report a weak but significant correlation between comprehension of quantifiers and Culture-Fair IQ ($r = 0.30$, $p < 0.05$) and between comprehension of

quantifiers and Need for Cognition ($r = 0.25$, $p < 0.05$);[7] and Street and Dąbrowska (2010, Experiment 2) found a moderately strong relationship between overall comprehension and Need for Cognition (*rho* = 0.576, $p < 0.001$). Perhaps most intriguingly, Skehan (1989) observed correlations of the order of 0.4 and above between scores on (foreign) language aptitude tests taken at age 13 and various measures of *first* language development at age 3. This raises the possibility that the HAA participants are simply better language learners – in other words, that they are better at inferring meaning from context, noticing patterns in the input, and generalizing those patterns – and that their academic success is a result of their better linguistic skills.

It should be pointed out that the various explanations offered above are not mutually exclusive. It is possible that less able individuals need more linguistic experience to achieve the same level of proficiency, and get less. Furthermore, different factors may contribute differently to knowledge of different constructions. Street and Dąbrowska (2010) provide some suggestive evidence that this might be the case: in their study, reading was a better predictor of performance on passives (which is not surprising, since the full passive is encountered primarily in formal written texts), while need for cognition was a better predictor of performance on quantifier constructions (which may be related to the fact that quantifiers play an important role in logical reasoning).

---

[7] Culture-Fair IQ is an intelligence test developed by Cattell and Cattell (1973) which is supposed to be culturally unbiased, or at least less biased than other IQ tests. The Need for Cognition test (Cacioppo et al., 1984) measures how much an individual enjoys effortful cognitive activity.

Conclusion

Numerous earlier studies have demonstrated substantial individual differences in sentence processing – in how fast people understand sentences, how quickly they recover from garden paths, and how they integrate lexical, syntactic and extralinguistic cues (King and Just 1991, Pearlmutter and Macdonald 1995, Swets et al. 2007; for a review, see Farmer, Misyak and Christiansen in press). Most of these studies tested comprehension of difficult structures such as object relatives and noun phrases with post-modifying prepositional phrases; the participants were usually undergraduate students, often at elite universities – i.e., highly literate. However, a recent ERP study by Pakulak and Neville (2010) found considerable differences between high and low proficiency participants (as assessed by a standardised language test) in the processing of very simple phrase structure violations: noun phrases containing an extra determiner such as *my his farm*. These studies are different from the research described here, in that the participants were able to process the sentences correctly (or, in the Pakulak and Neville study, detect the syntactic anomaly); thus, arguably, they had the relevant knowledge, and the observed differences could be described as differences in processing skill.

The results summarised in this paper, on the other hand, document differences in linguistic knowledge – i.e., competence. They thus contradict a widely held belief – that all first language learners (or at least, monolingual first language learners) converge on the same grammar. With the exception of some of the complex syntactic structures investigated by Dąbrowska (1997), all the studies described here deal with basic constructions that one

would not be surprised to hear in ordinary spoken discourse, including discourse addressed to young children, and which are believed to be acquired early in acquisition. Research on the acquisition of Polish case inflections shows that the genitive and dative endings emerge in the second year of life (Dąbrowska 2004, Smoczyńska 1985). The genitive endings are used productively from about 2;6 or even earlier, as evidenced by occasional overgeneralization errors (Dąbrowska 2001) and the ability to inflect nonce words in experimental settings (Dąbrowska and Szczerbiński 2006). Productivity with the dative endings develops somewhat later, but by 4;6, most Polish children are able to use at least the masculine and feminine endings productively, and some are also productive with the neuter ending (Dąbrowska and Szczerbiński 2006). English passives and universal quantifiers are somewhat later developments, although some researchers have suggested that they also develop in the preschool years. Pinker, Lebeaux and Frost (1987), for instance, found that children as young as four performed at ceiling on a comprehension task testing unfamiliar verbs in the passive voice, and many were also able to produce passives with novel verbs which they have only encountered in the active; and Crain et al. (1996) report excellent performance (88–98% correct responses) on tasks involving comprehension and production of sentences with universal quantifiers in children aged from 3;5 to 5;10.

How can we reconcile the acquisition research with the findings reported here? First, it is important to note that while a few researchers have observed excellent performance in very young children, the results reported in most studies are somewhat less impressive. For instance, Maratsos et al. (1985) and Gordon and Chafetz (1990) report chance (and, in some conditions, below chance) performance on passives in four- and five-

year-olds; and Geurts, in his overview of research on the comprehension of sentences with universal quantifiers observes that ‘‘error rates in excess of 50% are quite common’’ (2003:199) even in seven-year-olds. The differences between these studies could be due to a number of reasons, since they used different stimuli and different elicitation methods and tested different populations. With respect to the last point, it is important to note that most child language research is done with children from middle- to upper-middle-class backgrounds. There are good practical reasons for this: it is usually much easier to obtain parental consent from middle class parents, and the researcher can avoid complications due to the children acquiring a non-standard dialect. However, this means that most child language research is done with a very unrepresentative sample of the population: children from less privileged social backgrounds typically do less well as a group, and also show a great deal more individual variation (Ginsborg 2006, Huttenlocher et al. 2002, Locke and Ginsborg 2003). The problem is further exacerbated by the fact that some studies have very high drop-out rates (as high as 45% in some cases). Furthermore, many child language studies report considerable individual differences: for instance, in Dąbrowska and Szczerbiński’s (2006) study of the acquisition of Polish case inflections, individual performance in most age groups and conditions ranged from 0% to 100% correct. Thus, the apparent discrepancies between the adult studies reported here and some first language acquisition studies is due to the fact that there are considerable individual differences at all ages, and the fact that children of low academic achievement parents tend to be underrepresented in child language research. While some learners may know all there is to

know about a particular construction at age 4, some acquire this knowledge considerably later, and a small but significant minority may not acquire it at all.[8]

As pointed out in the introductory section, the assertion that all learners converge on more or less the same grammar is one of the strongest arguments for an innate language faculty, and one of the things that the Universal Grammar hypothesis was supposed to explain. The results summarised here suggest that the convergence argument is based on a false premise: there are, in fact, considerable differences in how much speakers know about some of the basic constructions of their native language. This does not necessarily mean that Universal Grammar does not exist: one can argue in favour of innate constraints on language learning on other grounds, for instance, poverty of the stimulus. It does, however, suggest that linguistic experience plays considerably more than a merely 'triggering' role in acquisition and that a substantial amount of experience with specific constructions (passives, noun complement clauses, etc.) is necessary to acquire mastery. Furthermore, the results of the training studies suggest that mere exposure does not guarantee acquisition: in

[8] It is interesting to note in this connection that the children in the Pinker et al. study, in which four-year-olds performed at ceiling on passive sentences, were recruited from day care centres affiliated with Harvard University, while those who participated in the studies with high error rates (e.g. Maratsos et al. 1985, Gordon and Chafetz 1990) came from more mixed backgrounds. For further discussion of the apparent discrepancies between the adult studies summarised here and first language acquisition research, and whether they can be attributed to differences in methodology, see Dąbrowska and Street (2006) and Street and Dąbrowska (2010).

order to acquire a particular structure, learners need to attend to both form and meaning, and this may be facilitated by explicit instruction.

The existence of individual differences also has important methodological implications for the language sciences. Most theoretical linguists rely on their own grammaticality judgments, or those of a few colleagues, for data. Psycholinguists generally use experimental methods, but the participants in their experiments are typically university students. Both groups of researchers tacitly assume that the conclusions reached on the basis of such data can be generalized to other groups of speakers. The findings reported here show that is not always the case. We cannot simply assume that what is true of one native speaker of a language will also be true of others: to make general statements about speakers of a particular language or language variety, we need to collect data from a range of speakers of different backgrounds.

It is true that more educated speakers' responses tend to be more consistent – that is to say, there is less variability both within and between participants. This is largely attributable to the fact that they tend to perform at ceiling: for instance, in the first experiment in the Street and Dąbrowska study, all the postgraduate students chose the correct picture 100% of the time in all conditions, while in the less educated group individual scores ranged from 0% to 100% on quantifier sentences and from 33% to 100% on passives. Thus, if we are interested in describing the linguistic abilities of educated speakers, even a very small sample may be perfectly adequate, at least for the basic constructions of the language. However, it is important to bear in mind that such speakers' linguistic abilities are shaped to a considerable extent by education and exposure to written language. In other words, in describing their knowledge, we are describing a *schooled*

linguistic competence, not competence in its natural state (cf. Perfetti and McCutchen 1987).

The existence of substantial individual differences in monolingual native speakers' knowledge of core grammatical constructions also has interesting implications for research on bilingualism and second language acquisition. A large number of studies comparing native and near-native L2 speakers have found age of acquisition effects (Birdsong 1992, 2004, 2006, Birdsong and Molis 2001, Coppieters 1987, Flege, Yeni-Komshian and Liu 1999, Johnson and Newport 1989): later L2 learners, particularly those who were exposed to the second language after puberty, typically diverge from the native speaker norm and are also more heterogeneous as a group – that is to say, they are more different from each other than the native speakers.

The research reviewed in this paper suggests that appeals to a native speaker norm are problematic because native speakers do not all converge on the same grammar. Moreover, while not all of the individual differences in native language attainment described here were related to education, for those that were, a very clear pattern emerged: the high academic achievement group consistently performed at ceiling, while the low achievement group showed much more variability, with some individuals performing at ceiling and some at chance, or even below chance, levels. It is interesting to observe in this context that the native control groups used in the ultimate attainment in L2 studies are nearly always university students or graduates – the native group that shows the highest attainment and the greatest amount of convergence. The divergent outcomes often observed in late bilinguals, then, may have more to do with schooling in the second language than age of acquisition per se. There is some evidence supporting such a conclusion. Flege et al.

(1999), for instance, in their study of 240 Korean-English bilinguals, found that age of arrival effects disappeared once the amount of schooling in the second language and the use of L2 were controlled for. Conversely, heritage speakers often fail to fully acquire their first (i.e., home) language if they do not receive schooling in it, even when the language is supported by a community of speakers outside the family (Montrul 2008).

A different, though related, issue is that of whether there are qualitative differences between L1 and L2 grammars. It has been suggested that L2 grammars may be incomplete, i.e. lack some properties of the grammar acquired by monolingual native speakers, or probabilistic rather than deterministic (see Birdsong 2004, Bley-Vroman 1990, 2009, Montrul 2008, Sorace 1993). Others (e.g. Herschensohn 2009) have argued that the differences are quantitative rather than qualitative, and may be more a matter of proficiency than age of acquisition. However, L1 grammars may also be incomplete or indeterminate (see also Montrul 2008, O'Grady, Lee and Lee in press). As we have seen, a substantial proportion of adult native speakers of English with relatively little schooling appear to be insensitive to quantifier scope, in that they do not distinguish between *Every fish is in a bowl* and *Every bowl has a fish in it.* Some of the other constructions discussed earlier – the Polish dative inflection, the English passive, and the complex sentence types studied by Dąbrowska (1997) – are arguably a part of the LAA participants' grammars, in that they show some degree of productivity in the case of the former, and typically perform above chance on tasks tapping comprehension of the latter two, but their performance is clearly well below ceiling. This also suggests that the differences in ultimate attainment are quantitative rather than qualitative – or, if they are qualitative, the split is not between native and non-native speakers but along some other dimension.

Another study which speaks to this issue is Dąbrowska and Street (2006), who, as we have seen, found that non-native speakers, including less educated non-natives, outperformed the LAA native group on a task tapping the comprehension of passive sentences. This result has been replicated and extended in another study which tested HAA and LAA non-native English speakers using the picture selection task from Street and Dąbrowska 2010 (Dąbrowska and Street in progress). It is unclear whether the surprisingly good performance of the LAA non-natives is due to the fact that they are bilinguals, that they have received explicit instruction in the second language, or some other factor; but the results indicate that L2 speakers can perform better on some grammatical tasks than monolinguals of a similar social and educational background.

Clearly, more research is needed before the full implications of individual differences in adult native speakers' grammatical knowledge can be assessed. We need to identify other constructions which may not be fully acquired by all first language learners – as well as constructions for which there is little, if any, individual variation; test more representative samples of the population to ascertain how widespread incomplete acquisition is; and establish which aspects of individual experience and/or cognitive and motivational factors are responsible for individual differences, and how the various causal factors interact with each other. Last but not least, given the relationship between linguistic abilities and academic achievement, it will be important to assess the impact of incomplete grammatical knowledge on individuals' lives.

References

Ambridge, B., Theakston, A. L., Lieven, E., & Tomasello, M. 2006. The Distributed Learning Effect for Children's Acquisition of an Abstract Grammatical Construction. *Cognitive Development, 21*, 74-193.

Bever, T. G. 1970. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 279-362). New York, NY: Wiley.

Bielec, D. 1998. *Polish: An Essential Grammar*. London: Rutledge.

Birdsong, D. 1992. Ultimate attainment in second language acquisition. *Language, 68*, 706-755.

Birdsong, D. 2004. Second language acquisition and ultimate attainment. In A. Davies & C. Elder (Eds.), *Handbook of Applied Linguistics* (pp. 82-105). London: Blackwell.

Birdsong, D. 2006. Age and second language acquisition and processing. A selective overview. *Language Learning, 56*, 9-46.

Birdsong, D., & Molis, M. 2001. On the evidence for maturational constraints in second language acquisition. *Journal of Memory and Language, 44*, 235-249.

Bley-Vroman, R. 1990. The logical problem of foreign language learning. *Linguistic Analysis, 20*, 3-49.

Bley-Vroman, R. 2009. The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31*, 175-198.

Brooks, P. J., & Sekerina, I. 2005/2006. Shortcuts to quantifier interpretation in children and adults. *Language Acquisition, 13*, 177-206.

Brooks, P. J., & Sekerina, I. A. 2006. Shallow processing of universal quantification: A comparison of monolingual and bilingual adults. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 2450.

Bybee, J. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425-455.

Bybee, J. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. 1984. The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.

Cattell, R. B., & Cattell, H. E. P. 1973. *Measuring Intelligence with the Culture-Fair Tests*. Champaign, IL: Institute for Personality and Ability Testing.

Chipere, N. 2001. Native speaker variations in syntactic competence: Implications for first language teaching. *Language Awareness, 10*, 107-124.

Coppieters, R. 1987. Competence differences between native and near-native speakers. *Language, 63*, 544-573.

Cowart, W. 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgements*. Thousand Oaks, CA: Sage Publications.

Crain, S., & Lillo-Martin, D. 1999. *An Introduction to Linguistic Theory and Language Acquisition*. Malden, MA: Blackwell.

Crain, S., Thornton, R., Boster, C., Conway, L., Lillo-Martin, D., & Woodams, E. 1996. Quantification without qualification. *Language Acquisition, 5*, 83-153.

Dąbrowska, E. 1997. The LAD goes to school: A cautionary tale for nativists. *Linguistics, 35*, 735-766.

Dąbrowska, E. 2001. Learning a morphological system without a default: The Polish genitive. *Journal of Child Language, 28*, 545-574.

Dąbrowska, E. 2001. Learning a morphological system without a default: The Polish genitive. *Journal of Child Language, 28*, 545-574.

Dąbrowska, E. 2004. *Language, Mind and Brain. Some Psychological and Neurological Constraints on Theories of Grammar*. Edinburgh: Edinburgh University Press.

Dąbrowska, E. 2008a. The later development of an early-emerging system: The curious case of the Polish genitive. *Linguistics, 46*, 629-650.

Dąbrowska, E. 2008b. The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language, 58*, 931-951.

Dąbrowska, E. 2009. Constructing a second language: Some final thoughts. *Annual Review of Cognitive Linguistics, 7*, 277-290.

Dąbrowska, E. and Street, J. Submitted. When L2 learners outperform native speakers.

Dąbrowska, E., & Street, J. 2006. Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences, 28*, 604-615.

Dąbrowska, E., & Szczerbiński, M. 2006. Polish children's productivity with case marking: the role of regularity, type frequency, and phonological coherence. *Journal of Child Language, 33*, 559-597.

Farmer, T. A., Misyak, J. B., & Christiansen, M. H. In press. Individual differences in sentence processing. In M. Spivey, M. Joannisse & K. McRae (Eds.), *Cambridge Handbook of Psycholinguistics*. Cambridge: Cambridge University Press.

Ferreira, F. 2003. The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*, 164-203.

Ferreira, F., & Patson, N. D. 2007. The 'good enough' approach to language comprehension. *Language and Linguistics Compass, 1*, 71-83.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11-15.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. 1999. Age constraints on second-language acquisition. *Journal of Memory and Language, 41*, 78-104.

Gathercole, S., Pickering, S., Knight, C., & Stegmann, Z. 2004. Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology, 18*, 1-16.

Geurts, B. 2003. Quantifying kids. *Language Acquisition, 11*, 197-218.

Ginsborg, J. 2006. The effects of socio-economic status on children's language acquisition and use. In J. Clegg & J. Ginsborg (Eds.), *Language and Social Disadvantage* (pp. 9-27). Chichester: John Wiley & Sons.

Gordon, P., & Chafetz, J. 1990. Verb-based versus class-based accounts of actionality effects in children's comprehension of passives. *Cognition, 36*, 227-254.

Gries, S. Th., & Stefanowitsch, A., 2004. Extending collostructional analysis. A corpus-based perspective on ‘alternations’. *International Journal of Corpus Linguistics* 9, 97–129.

Hart, B., & Risley, R. R. 1999. *The Social World of Children Learning to Talk*. Baltimore: Paul Brookes.

Hart, B., & Risley, T. R. 1995. *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: Paul Brooks.

Herschensohn, J. 2009. Fundamental and gradient differences in language development. *Studies in Second Language Acquisition, 31*, 259-289.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. 2002. Language input and child syntax. *Cognitive Psychology, 45*, 337-374.

Johnson, J., & Newport, E. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60-99.

Just, M. A., & Carpenter, P. A. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

King, J., & Just, M. A. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580-602.

King, J., & Just, M. A. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580-602.

Lidz, J., & Williams, A. 2009. Constructions on holiday. *Cognitive Linguistics, 20*, 177-189.

Locke, A., & Ginsborg, J. 2003. Spoken language in the early years: the cognitive and linguistic development of three- to five-year-old children from socio-economically deprived backgrounds. *Educational and Child Psychology, 20*, 68-79.

MacDonald, M. C., & Christiansen, M. H. 2002. Reassessing Working Memory: Comment on Just and Carpenter (1992) and Waters and Caplan 1996. *Psychological Review, 109*, 35-54.

Maratsos, M., Fox, D. E. C., Becker, J. A., & Chalkley, M. A. 1985. Some restrictions on children's passives. *Cognition, 19*, 167-191.

Montrul, S. 2008. *Incomplete Acquisition in Bilingualism: Re-Examining the Age Factor*. Amsterdam: John Benjamins.

Nowak, M. A., Komarova, N. L., & Niyogi, P. 2001. Evolution of Universal Grammar. *Science, 291*, 114 - 118.

O'Grady, W., Lee, O.-S., & Lee, J.-H. In press. Practical and theoretical issues in the study of language acquisition. *Heritage Language Journal*.

Orzechowska, A. 1998. Rzeczownik. In R. Grzegorczykowa, R. Laskowski & H. Wróbel (Eds.), *Gramatyka współczesnego języka polskiego: Morfologia* (2nd ed., Vol. 1, pp. 270-332). Warszawa: PWN.

Otwinowska-Kaszelanic, A. 2000. *Korpus języka mówionege młodego pokolenia Polaków*. Warszawa: Wydawnictwo Akademickie DIALOG.

Pearlmutter, N. J., & MacDonald, M. C. 1995. Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language, 34*, 521-542.

Perfetti, C. A., & McCutchen, D. 1987. Schooled language competence: linguistic abilities in reading and writing. In S. Rosenberg (Ed.), *Advances in Applied Psycholinguistics. Volume 2: Reading, Writing, and Language Learning* (pp. 105-141). Cambridge: Cambridge University Press.

Pinker, S., Lebeaux, D. S., & Frost, L. A. 1987. Productivity and constraints in the acquisition of the passive. *Cognition, 26*, 195-267.

Ross, J. R. 1979. Where's English? In C. J. Fillmore, D. Kempler & W. S.-Y. Wang (Eds.), *Individual Differences in Language Ability and Language Behavior* (pp. 127-163). New York: Academic Press.

Schmidt, R. W. 1990. The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158.

Schütze, C. T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.

Seidenberg, M. S. 1997. Language acquisition and use: learning and applying probabilistic constraints. *Science, 275*, 1599-1603.

Skehan, P. 1989. *Individual Differences in Second-Language Learning*. London: Edward Arnold.

Smoczyńska, M. 1985. The acquisition of Polish. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition. Volume 1: The Data* (pp. 595-683). Hillsdale, NJ: Lawrence Erlbaum.

Sorace, A. 1993. Incomplete vs divergent representations of unaccusativity in non-native grammars of Italian. *Second Language Research, 9*, 22–47.

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. Submitted. Lexically specific knowledge and individual differences in adult native speakers' processing of the English passive.

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. 2007. The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*, 64-81.

Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Child Language Acquisition*. Cambridge, MA: Harvard University Press.

Townsend, D. J., & Bever, T. G. 2001. *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, MA: MIT Press.

Waters, G. S., & Caplan, D. 1996. The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, *49*, 51-79.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250-271.