Lexically-specific knowledge and individual differences in adult native speakers’ processing of the English passive

James A. Street
Department of Humanities, School of Arts and Social Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST james.street@northumbria.ac.uk
Tel: +44 (0) 191 243 7310
Fax: +44 (0) 191 227 3696

Ewa Dąbrowska
Department of Humanities, School of Arts and Social Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST ewa.dabrowska@northumbria.ac.uk
Tel: +44 (0) 191 227 3497
Fax: +44 (0) 191 227 3696
Lexically-specific knowledge and individual differences in adult native speakers’ processing of the English passive

Abstract

This paper provides experimental evidence for the role of lexically specific representations in the processing of passive sentences and considerable education-related differences in comprehension of the passive construction. The experiment measured high academic attainment (hereafter, HAA) and low academic attainment (hereafter, LAA) participants’ response time and decision accuracy using an online task which compared processing and comprehension of active and passive sentences containing verbs strongly associated with the passive and active constructions, as determined by collostructional analysis (see e.g., Gries & Stefanowitsch, 2004; Stefanowitsch & Gries, 2003). As predicted by usage-based accounts, participants’ performance was influenced by frequency (both groups processed actives faster than passives; the LAA participants also made significantly more errors on passive sentences) and lexical specificity (i.e., processing of passives was slower with verbs strongly associated with the active). Contra to proposals made by Dąbrowska & Street (2006), the results suggest that both HAA and LAA participants have verb-specific as well as verb-general representations, but that the latter are less well entrenched in LAA participants, resulting in less reliable performance. The results also show no evidence of a speed-accuracy trade-off, making alternative accounts of the results (e.g., those of two-stage processing models such as Townsend & Bever, 2001) problematic.
Keywords: Lexically specific knowledge, Collostructional analysis, Individual differences, English passive, Sentence comprehension, Sentence processing, Linguistic competence, Linguistic performance, Usage-based, LAST
1. Introduction

Most linguists assume, either implicitly or explicitly, that all first language learners converge on the same grammar (see e.g., Birdsong, 2004: 83; Bley-Vroman, 2009: 179; Chomsky, 1975: 11; Crain & Lillo-Martin, 1999: 9; Lidz & Williams 2009: 177; Nowak et al., 2001: 114; Seidenberg, 1997: 1600). Although it is well established that there are large individual differences in lexical knowledge and knowledge of archaic, very formal or highly literary grammatical constructions (e.g., *Little did I know that...*), all speakers are believed to share the same ‘core’ grammar. Linguists who assume convergence rarely justify it in any way: it is regarded as a self-evident, incontrovertible “fact” which linguistic theory must explain. Indeed, along with poverty of the stimulus, convergence is regarded as an incontestable argument for an innate universal grammar: if different learners are exposed to different input yet acquire essentially the same grammar, the construction of grammar must be guided by highly restrictive principles (cf. Chomsky, 1975, Lidz & Williams, 2009). Conversely, the fact that second language learners typically do not converge has been used to argue that L2 learning is fundamentally different from first language acquisition (cf. Bley-Vroman, 1990, 2009).

However, there is a growing body of evidence which suggests that there are considerable differences in native speakers’ mastery of some basic grammatical constructions. Dąbrowska (2008a), for example, reports that different groups of Polish speakers appear to have learned different generalizations about the distribution of genitive masculine inflections: some rely on very general semantic cues (animacy), others
use more specific semantic information (e.g., substance v. object), and still others are sensitive primarily to phonological properties of the stem. These differences are not explainable in terms of regional or social variation, although they are not, perhaps, very surprising, as this part of the inflection system is highly irregular, and therefore it is not clear what the “correct” generalization would be. However, Dąbrowska (2008b) found large individual differences with the Polish dative inflection, which is almost completely regular. In a nonce word inflection experiment participants were asked to supply the dative form of masculine, feminine and neuter nouns. Within each gender, half the nouns belonged to densely populated neighbourhoods (i.e., they resembled many existing nouns) and half to sparsely populated neighbourhoods. Dąbrowska found that participants reliably supplied the target forms of nonce nouns belonging to large classes and densely populated neighbourhoods, indicating that they were cooperative and had understood the task. However, results for nouns belonging to smaller classes and/or sparsely populated neighbourhoods were much more varied, with individual scores ranging from 29% to 100% (4%-100% for nouns from low-density neighbourhoods). Moreover, performance on the inflection task was strongly correlated with number of years in formal education, with more educated participants achieving higher scores. A smaller-scale study by Wolff (1981) reports a similar finding for German participles.

There is also evidence of considerable variation in native speakers’ ability to understand complex syntactic constructions. Dąbrowska (1997) presented participants of varying educational backgrounds with complex syntactic structures such as the tough movement construction (e.g., John will be hard to get his wife to vouch for), sentences with parasitic gaps (e.g., It was King Louis who the general convinced that this slave
might speak to) and complex NPs (e.g., Paul noticed that the fact that the room was tidy surprised Shona). Comprehension was tested using simple questions about the sentences (e.g., What did Paul notice? What surprised Shona? for the complex NPs). Dąbrowska found a strong relationship between educational attainment and performance on the comprehension task: university lecturers outperformed graduate students who in turn were better than undergraduates who were better than unskilled workers with no university education.

The complex syntactic structures used in this study place heavy processing demands on participants and, therefore, the less educated participants’ failure to respond correctly may be attributable to working memory limitations rather than being a reflection of their linguistic competence. However, the results reported in Chipere (2001, 2003) suggest that this is not the case. Chipere tested high academic attainment (hereafter, HAA) and low academic attainment (LAA) participants’ ability to comprehend and recall complex NP sentences. In the first phase of the experiment, the LAA group performed significantly worse than the HAA group on both tasks. In the second phase Chipere divided the LAA participants into two subgroups which were given different types of training. One group was given memory training (i.e., participants were asked to repeat complex NP sentences). The other group was given comprehension training (i.e., participants received explicit instruction about complex NP constructions followed by a practice session with feedback). Both groups were then tested again with new complex NP sentences. Chipere found that memory training resulted in improved performance on the recall task, but not the comprehension task, whilst comprehension training led to an improvement in performance on both tasks. These results suggest that the LAA group’s
poor performance on the initial comprehension test was not attributable to limitations in working memory capacity, but rather to lack of experience with the complex NP construction.

Similar conclusions can be drawn from another training study conducted by Wells et al. (2009). Wells et al. point out that people with low reading spans have particular difficulty with object relatives such as [The reporter that the senator attacked] admitted the error. This is traditionally attributed to low working memory capacity, which has a disproportionate effect on structures that place heavy demands on working memory (Just & Carpenter, 1992). Wells et al., following MacDonald and Christiansen (2002), suggest that the differences might instead be attributable to differences in linguistic experience: people who read more have more experience with language, particularly with more difficult structures, and hence are faster and more accurate at interpreting object relatives; they also perform better on the reading span task because the task actually measures language processing skill, not working memory capacity. In their study, Wells et al. exposed undergraduate students to 160 sentences containing subject and object relatives in two training sessions. A post-test administered four days after the second training session revealed that reading times at the main verbs for object relatives decreased as a result of training, while there was no analogous effect in the control group who had been exposed to different types of sentences. This is consistent with the idea that the individual differences in the processing of object relatives observed in earlier studies are attributable to differences in the amount of experience with this structure.

Several recent studies (e.g., Dąbrowska & Street 2006; Street 2009; Street & Dąbrowska 2010) have suggested that there are considerable education-related
differences in speakers’ mastery of syntactically less complex constructions such as the English passive. The passive, whilst undeniably part of ‘core’ grammar, does not involve embedding and thus can be plausibly assumed not to place such a heavy burden on working memory as the constructions used by Dąbrowska (1997), Chipere (2001) and Wells et al. (2009). Furthermore, since the full passive occurs predominantly in formal written texts,\(^1\) more educated speakers (who tend to read more) have considerably more experience of this construction. According to usage-based accounts (e.g., Bybee 2010, Langacker 2000, MacDonald and Christiansen 2002), more experience with a construction should result in greater entrenchment and hence improved performance. Dąbrowska & Street (2006) tested this prediction employing a modified version of a task developed by Ferreira (2003) in which participants were asked to identify the ‘do-er’ (i.e., AGENT) of plausible and implausible reversible passive sentences (e.g., *The man was bitten by the dog, The dog was bitten by the man*). As anticipated, the HAA group performed at ceiling, while the LAA participants had considerable problems with implausible passives, suggesting that they processed the test sentences non-syntactically (i.e., relied on world knowledge rather than grammatical knowledge).

Further evidence for individual differences in grammatical knowledge comes from research by Street (2009) and Street & Dąbrowska (2010), who tested three structures: reversible passives (e.g., *The girl was kissed by the boy*), locative sentences with the universal quantifier *every* (e.g., *Every fish is in a bowl*; hereafter Q-*is*), and possessive locatives with *every* (e.g., *Every bowl has a fish in it*; hereafter, Q-*has*). In addition, reversible actives (*The boy kissed the girl*) were used as a control condition. The first experiment tested HAA (i.e., postgraduate) and LAA (i.e., nongraduate) native English
speakers using a picture selection task. 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 (i.e., three bowls containing a fish plus one empty bowl, and three bowls containing a fish plus one fish without a bowl). The results of Experiment 1 revealed that whilst the graduate participants performed at ceiling in all conditions, the nongraduates performed at ceiling only on the actives (97% correct). Their performance was somewhat worse on passives (88%), worse still on Q-is sentences (78%), and at chance on Q-has (43%). Interestingly, group performance reflected order of frequency of the respective constructions in the British National Corpus (i.e., active > passive > Q-is > Q-has). This could be due to differences in entrenchment, although, since different constructions are involved, other interpretations are also possible.

The second experiment was a training study, in which adult literacy students were tested before and after training on sentences similar to those used in Experiment 1. A pre-test was used to select low-scoring participants who were then randomly assigned to either a quantifier training group or passive training group. Training comprised a short “grammar lesson” on either the passive or the quantifier construction, followed by a practice session with feedback. Immediately after training, participants were tested again with a different version of the task. To see if any effects of training were long-lasting, participants were retested 1 week and 12 weeks after training, again with different versions of the task.

The pre-test results were very similar to those of the LAA group in Experiment 1, with order of difficulty reflecting the relative frequency of the constructions. The post-
training results show that training resulted in selective improvement. Participants trained on the quantifier constructions showed a clear improvement in performance on the quantifier conditions (where the mean scores rose from 37% in the Q-is condition and 13% in the Q-has condition to 94% and 100% respectively on the first post-test), but not on the passive, whilst participants trained on the passive showed a clear improvement in performance on those sentences (from 48% and to 98%), but not on the quantifier sentences. This pattern of results was observed in all the post-training tests, showing that the effects of training were long-lasting. Thus, performance improved dramatically as a result of additional experience with the relevant construction, indicating that the initial differences in test scores are attributable to differences in specific linguistic knowledge.

The differences observed in Dąbrowska & Street (2006) and Street & Dąbrowska (2010) indicate that some native speakers have problems in interpreting passive sentences, and that the differences between speakers are at least partly related to educational level: the HAA participants’ performance was invariably at ceiling, whilst a vast amount of variation was observed in the performance of LAA individuals. Indeed, with particular regard to the LAA group’s performance on passive constructions, such variation was observed with different types of passive sentences (i.e., implausible as well as unbiased) and across different testing paradigms (i.e., naming the ‘do-er’ and picture selection).

How can we account for these education-related differences in the interpretation of passive sentences? One possibility is that the low-performing speakers have in fact mastered this construction, and their inability to reliably provide the correct answer in an experimental situation is attributable to performance problems. It is clear that the findings
summarised above cannot be explained away by appealing to linguistically irrelevant
performance factors such as failure to attend or cooperate with the researcher or inability
to understand instructions, since the same participants performed at ceiling on active
sentences. They also cannot be attributed to working memory limitations, if working
memory is conceptualised as a resource with a fixed capacity, since in the second
experiment reported by Street & Dąbrowska performance improved dramatically
following a training session lasting only about 5 minutes. However, it is possible that the
observed differences in performance are attributable to individual differences in parsing
ability rather than linguistic knowledge. Specifically, the results could be accommodated
by a two-stage processing theory such as LAST (Late Assignment of Syntax Theory; see
Townsend and Bever, 2001) according to which sentence processing involves two
distinct phases. In the first phase, the processing system constructs a ‘‘pseudoparse’’, a
rough analysis based on superficial probabilistic cues and heuristics such as the so-called
NVN strategy for assigning thematic roles to predicates (see Bever, 1970, Townsend and
Bever, 2001).2 The pseudoparse is then used to guide the true parse, an algorithmic
process which accesses syntactic knowledge to construct a complete syntactic
representation. Constructing the true parse is slower and computationally more
demanding, and thus may not be carried out in certain circumstances (e.g., under time
pressure, or when processing resources are limited). Under this approach, the LAA
participants would be assumed to have the same grammatical knowledge as the HAA
group, but be less likely to perform the true parse.

An alternative explanation, suggested by Dąbrowska & Street (2006), is that LAA
participants lack a well-entrenched verb-general passive construction and therefore rely
predominantly on lexically specific representations for individual verbs. Lexically specific representations are known to play an important role in language acquisition. Young children’s’ syntactic representations have been shown to be tied to particular lexical items, typically verbs (Tomasello, 2003, Abbot-Smith and Tomasello, 2006, Behrens 2009). Later in development, children acquire more abstract units by generalizing over the early lexically specific patterns. However, as demonstrated by the large body of work in the constraint-based lexicalist framework (Garnsey et al., 1997, Holmes et al., 1989, MacDonald, Perlmutter & Seidenberg, 1994, MacDonald & Seidenberg, 2006, Trueswell et al., 1993, Trueswell & Tanenhaus, 1994), adults are very sensitive to the distributional peculiarities of individual words, suggesting that lexically-specific representations survive into adulthood.

More abstract patterns are more difficult to acquire and process (Dąbrowska 2010), and hence their full mastery requires exposure to a relatively large number of exemplars. Since passive sentences occur predominantly in formal written language, less educated speakers have comparatively little experience of this construction, and may develop only a weak passive schema which cannot be reliably accessed during language use, or fail to develop a verb-general passive construction at all. Such individuals would still be able to produce and understand some passive sentences by relying on verb-specific representations \((NP1 \text{ was known by } NP2, \ NP1 \text{ is based on } NP2, \text{ etc.})\) for verbs which frequently occur in the passive. However, if they encountered a passive sentence for which they lacked a verb-specific schema, they would have to either rely on analogy, which is costly in processing terms, or process it nonsyntactically (i.e., using a processing heuristic such as the NVN strategy), which leads to errors (cf. Dąbrowska & Street,
2006). Thus, according to this account, inconsistent individual performance on passive sentences is attributable to a lack of a well-entrenched verb-general passive construction.

This paper examines the role of lexically specific representations and processing heuristics in the comprehension of passives by LAA and HAA adults. We use a version of a task developed by Ferreira (2003) in which participants are presented with active and passive sentences and asked to identify the “do-er” (i.e., AGENT) or the “acted-on” (i.e., the PATIENT). Ferreira’s study compared performance on plausible passives such as *The man was bitten by the dog*, symmetrical passives such as *The man was visited by the woman*, and implausible passives like *The dog was bitten by the man*. The participants (undergraduate students) performed above chance in all conditions, indicating that they had the relevant syntactic knowledge. However, error rates were relatively high: about 12% for plausible sentences, 18% for symmetrical sentences, and 23% for implausible sentences. This, according to Ferreira, indicates that they did not always compute the full parse, but often relied on a “quick and dirty” pseudoparse – for instance, assumed that NVN = AGENT VERB PATIENT.

The aim of our study is two-fold: (1) to determine whether the education-related differences observed in earlier studies can be attributed to differences in the use of processing heuristics, specifically, the fact that some participants may not compute a full parse; and (2) to evaluate the role of lexically specific representations in the processing of passive sentences by HAA and LAA participants. To do this we compare participants’ comprehension of passive sentences containing verbs which are strongly associated with active sentences (hereafter, PVₐ) and verbs which are strongly associated with passives (hereafter, PVᵢ). We also compare participants’ comprehension of active sentences
containing verbs strongly associated with passive constructions (AV$_P$) and verbs strongly associated with active constructions (AV$_A$). The active sentences serve as a control condition: all speakers should have a well-entrenched verb-general active construction, and therefore no education-related differences are expected.

2 Predictions

Our experiment was designed to test two usage-based predictions and an additional prediction derived from the Late Assignment of Syntax Theory. According to usage-based theories (Bybee, 2010, Langacker, 2000), language use is influenced by two main factors: frequency and lexical specificity. Repeated experience with a particular construction leads to greater entrenchment, which in turn results in faster and more accurate processing. Therefore we expect active sentences to be easier than passive sentences for all participants. Moreover, since full passives occur predominantly in formal written texts, differences between active and passive sentences should be particularly pronounced in less educated participants, who have relatively little experience of such texts. With regard to lexical specificity, usage-based theories maintain that more abstract (i.e., lexically underspecified) constructions are more difficult to access. Thus, all other things being equal, participants should do better (i.e., process more quickly and interpret more accurately) in those sentences which contain verbs for which they have lexically specific representations – i.e., do better in passive sentences containing verbs strongly associated with the passive (PV$_P$) than in passive sentences containing verbs strongly associated with the active (PV$_A$). An analogous difference between active-attracting and passive-attracting verbs in active sentences would also be
compatible with a usage based model; however, since active sentences are known to be processed very quickly and accurately, it is likely that these differences will be masked by ceiling effects. Thus the predicted order of difficulty is that active-attracting verbs used in active sentences ($AV_A$) and passive-attracting verbs used in active sentences ($AV_P$) will be easier than passive-attracting verbs in passive sentences ($PV_P$) which, in turn, will be easier than active-attracting verbs in passive sentences ($PV_A$):

Prediction 1: $AV_A, AV_P > PV_P > PV_A$

We expect that the HAA participants will perform at ceiling, so the differences will only show up in reaction time data. For the LAA group, the effects are likely to be visible on both measures.

Furthermore, if, as hypothesised by Dąbrowska & Street (2006), LAA participants rely more on lexically specific knowledge about passives, lexical effects should be more pronounced for the LAA group – in other words, there should be an interaction between construction, verb and group.

Prediction 2: For the HAA group, $PV_P > PV_A$; for the LAA group, $PV_P >> PV_A$

Two-stage processing theories such as LAST do not predict differences between active- and passive-attracting verbs in either construction, since the same general rules and parsing routines are thought to apply in both cases. However, they do predict a speed-accuracy trade-off for passives: since the true parse requires additional processing time, participants who are more accurate should respond more slowly than the less accurate participants.

Prediction 3: There should be a positive correlation between accuracy and reaction time for passives.
3 Method

We employed a 2x2x2 mixed design with one between-group independent variable: level of academic attainment (HAA v. LAA) and two within-group independent variables: construction (active v. passive) and verb (active-attracting v. passive-attracting). The dependent variables are decision accuracy and response time. Participants read simple active transitive and full passive sentences and at the end of each identified a named participant in the sentence as either the ‘do-er’ (i.e., AGENT) or the ‘acted on’ (i.e., PATIENT; see Ferreira, 2003 for a similar naming task). The dependent variables were decision accuracy and reaction times.

3.1 Participants

Sixty-four adults (25 males and 39 females) aged 17–50 participated in the experiment. The HAA group comprised thirty-one postgraduate students or recent graduates. All participants had at least an MA degree (and hence at least 17 years of formal education); some were studying for or had recently completed a PhD. They came from a variety of academic disciplines (arts and humanities, social sciences and life sciences). The remaining thirty-three participants (the non-graduate, LAA group) had had at most 11 years of formal education and were employed as packers, cleaners and hairdressers. All participants were native speakers of English. One LAA participant did not fully engage
with the task and talked throughout the trials and was therefore excluded from the data analyses. Thus, the final sample comprised sixty-three participants.

3.2 Materials

There were 24 test sentences containing 12 verbs. Six of the verbs were associated with the active transitive construction; the other six were associated with the passive (based on the ICE-GB corpus). To determine which verbs are strongly associated with active and passive constructions we employed collostructional analysis (Stefanowitsch and Gries 2003, Gries and Stefanowitsch 2004). Collostructional analysis measures the association of words to syntactic constructions rather than words to words (as in traditional collocational analysis). This involves two steps. First, the analyst establishes the frequency of the target word in the target construction (e.g. *kick* in the passive), the target word in a contrasting construction (i.e., *kick* in the active), the frequency of all other words in the target construction (i.e., all verbs other than *kick* in the passive), and the frequency of all other words in the contrasting construction (in this case, all other verbs in the active). Then the observed frequency of the word in the construction is compared to one expected by chance, given the frequency of the verb and the frequency of the other constructions. The strength of attraction or repulsion of the word to the construction is measured using the $p$-value of a Fisher-Yates exact test which in addition to incorporating the size of the effect (normally obtained using $\eta^2$, $d$ or $r^2$) also weighs the effect on the basis of the observed frequencies and has the added advantage of not
overestimating the association of strength of infrequent data (see Stefanowitsch and Gries 2003 pp. 218-219 for further discussion).

Many of the most distinctive collexemes (i.e., lexemes which can occur in either the active transitive or passive but appear in one of the constructions significantly more often than would be expected by chance) of the two constructions are psychological verbs such as hate, impress, and remember. However, pilot studies revealed that a number of participants were unsure about assigning the labels ‘do-er’ and ‘acted-on’ to participants in sentences containing such verbs; and some participants were inclined to identify the PATIENT as the ‘do-er’ in some cases. For example, in a sentence such as Daniel hated Peter, some participants thought that Peter was the ‘do-er’ of some action which had caused Daniel to hate him. To avoid such ambiguities, all the verbs used in the experiment designated dynamic transitive events which readily allow two animate arguments. As a consequence, whilst the verbs are distinct collexemes for their respective constructions, they are not the most distinct collexemes for these constructions. A list of all the verbs used in the experiment, their frequencies in the active and passive in the ICE-GB corpus, and the Fisher-Yates $p$-value computed on the basis of these figures are given in Table 1. Note that since attraction to a construction is calculated relative to the construction’s frequency, and actives are much more frequent than passives, verbs attracted to the passive are not necessarily more frequent in the passive. Since the absolute frequency of these verbs in the ICE-GB is relatively small, we have also included estimates of the frequency of these verbs in each construction in the British National Corpus. These were computed on the basis of a random sample of 100 tokens of each verb.
Each verb occurred in both the active and passive construction. Thus of the 24 test sentences, six were active transitive containing an active-attracting verb ($V_A$), six were active transitive containing a passive-attracting verb ($V_P$), six were passives containing a $V_A$, and six were passives containing a $V_P$. Twelve pairs of names (six male and six female) were used for the AGENT and PATIENT thematic roles in the test sentences. Twelve sentences contained male names and the other twelve sentences contained female names. Thus for each of the 24 test sentences the NPs occupying the AGENT and PATIENT roles were of the same gender. These were used instead of common nouns in order to reduce noise by eliminating the effects of pragmatics (e.g., participants may feel that a man is more likely to attack a woman than vice versa). Each name within each pair appears 4 times: once as AGENT of an active sentence; once as AGENT of a passive sentence; once as PATIENT in an active transitive; and once as PATIENT in a passive. Since passive sentences contain additional morphological markers (the auxiliary be and the preposition by), they were slightly longer than the actives; however, within the same construction, all the sentences were of approximately the same length (i.e., contained approximately the same number of letters and same number of syllables). Example sentences of each type are given in Table 2.

There were four versions of the test, each containing six sentences for each of the four conditions, and within any one version there were no repeats of the same action involving the same participants (i.e., in any one version no name appears with the same verb twice). For descriptions of simple transitive events there are four possible descriptions (e.g., Mary punched Sandra, Sandra punched Mary, Mary was punched by Sandra and Sandra was punched by Mary); each of the four possible descriptions
appeared in a different version of the test. Ultimately, each participant saw 24 sentences: 6 AVₐ, 6 AVₚ, 6 PVₚ and 6 PVₐ with no repeats of the same names with the same verbs and each name appearing twice in the AGENT role and twice in the PATIENT role. The order of sentences was randomized for each participant. A complete list of sentences used in one version of the test is given in Appendix I.

3.3 Procedure

The experimental session began with participants reading written instructions displayed on a laptop screen. Participants were informed that they would be presented with a series of sentences followed by a name, for instance:

Justin scratched Steven

Justin

and that their task was to identify the person named as either the “do-er” (i.e., AGENT) or the “acted-on” (i.e., PATIENT). This was followed by a brief non-technical explanation of the terms “do-er” and “acted-on”. A complete transcript of the written instructions is given in Appendix II.

Participants responded by pressing ‘D’ (for ‘do-er’) or ‘A’ (for ‘acted-on’) on the keypad. Both these keys were highlighted (using white stickers situated beneath the letter at the bottom of the key). The sentence remained on screen until a participant pressed
either ‘D’ or ‘A’; it then disappeared and was replaced by a short sign (++) and shortly afterwards, the next sentence.

Before the test trials began, all instructions were clarified (and concept checked) verbally by the experimenter. Participants then completed four practice trials (two active, two passive). These were supervised by the experimenter to ensure that participants had understood the task. Participants were tested individually, with each testing session lasting approximately 5 minutes. The stimuli were presented using E-prime software (Psychology Software Tool, Pittsburgh, PA), which also recorded the participants’ decision accuracy and reaction times.

4. Results

4.1 Decision accuracy

The accuracy data are summarised in Table 3 below. As can be seen from the table, the HAA group performed at ceiling in all conditions. The LAA group also performed at ceiling on active sentences, but had considerably lower scores for passives. As the results are not normally distributed, the data were analysed using non-parametric tests (Kruskal-Wallis). These show that there are no significant differences between groups on decision accuracy for the active sentences: for AV_A, (H(1) = 1.017, p = .313); for AV_P, (H(1) = .153, p = .696). However, there are significant differences between groups in decision accuracy for the passives: for PV_A, (H(1) = 4.587, p = .032); for PV_P, (H(1) = 9.843, p = .002).
In order to further examine the effects of construction and the effects of verb we conducted two sets of planned comparisons for each group. To examine the effect of construction we compared participants’ scores for active and passive sentences with each type of verb (i.e., AVA vs. PV, and AVP vs. PV). To examine the effect of verb type we compared participants’ scores for active-attracting and passive-attracting verbs in each construction (i.e., AVA vs. AVP and PV vs. PV). Again, as the results are not normally distributed, the scores were compared using a non-parametric equivalent of the paired t-test: the Wilcoxon Signed Ranks. Comparisons adjusted using the Bonferroni correction give a probability value of $p = .016$.

4.1.1 Effect of construction

At this revised level the results show that for the HAA group there is a small but significant difference between AVA and PV: $z = -2.530; p = .011, N = 31, r = 0.22$. However, there is no significant difference between AVP and PV: $z = 1.000; p = 1.000, N = 31, r = 0.08$. By contrast, Wilcoxon tests for the LAA group reveal that both comparisons are highly significant at the corrected probability level of $p = .016$, and that the difference is much greater than that of the HAA group: for AVA vs. PV, $z = -3.384; p = .001, N = 32, r = 0.29$; for AVP vs. PV, $z = -2.810; p = .005, N = 32, r = 0.25$. Thus, the LAA group made more errors on both types of passive sentences. This is also evident if results for the two verb types are collapsed (cf. active and passive columns in Table 3). Wilcoxon tests reveal that whilst differences in decision accuracy for actives and passives
is not significant for the HAA group \((z = -1.890; \ p = .059, \ N = 31, \ r = 0.16)\), they are highly significant for the LAA participants \((z = 3.532; \ p < .001, \ N = 32, \ r = 0.31)\).

4.1.2 Effect of verb type

With regard to the effect of verb type, as anticipated, there was no significant difference between \(AV_A\) and \(AV_P\) for either group: for the HAA participants, \(z = -1.342; \ p = .180, \ N = 31, \ r = 0.12\); for the LAA participants, \(z = 2.000; \ p = 1.000, \ N = 32, \ r = 0.17\). However, the predicted difference between \(PV_A\) and \(PV_P\) was not observed: for HAA participants, \(z = -1.890; \ p = .059, \ N = 31, \ r = 0.16\); and for the LAA group, \(z = .053; \ p = .958, \ N = 32, \ r = 0\).

4.1.3 Individual differences

As can be seen from the standard deviations and ranges in Table 3, there were considerable individual differences in the LAA participants’ performance on passive sentences. There were a total of 12 passive sentences on the test. According to the binomial distribution \((p < .05)\), above chance performance requires 10 out of 12 correct responses, and a score of 2 or less would be below chance. At this criterion, 24 of the LAA participants, i.e. 75%, performed above chance; 7 were at chance (scores between 3 and 10), and one participant performed below chance, supplying the correct response on only one of the passive trials. (This participant responded correctly to 11/12 actives, so
s/he consistently interpreted the first NP in the sentence as the agent). All of the HAA participants performed above chance.

4.2 Response Time

Mean response times and SDs for all conditions by groups are summarised in Table 4. The data were analysed by means of a 2 (construction: active versus passive) by 2 (verb: active-attracting versus passive-attracting) by 2 (group: HAA versus LAA) mixed-design factorial ANOVA. The dependent measure was response time. All effects are reported as significant at $p < .05$. The analysis revealed a main effect of construction ($F_1(1,61) = 111.681, p < .001$, partial $\eta^2 = .647$; $F_2(1,10) = 118.362, p < .001$, partial $\eta^2 = .922$) indicating that, as expected, both groups were faster at processing active constructions than passive constructions. There was also a main effect for group ($F_1(1,61) = 7.913, p < .007$, partial $\eta^2 = .115$; $F_2(1,10) = 61.644, p < .001$, partial $\eta^2 = .860$), indicating that the HAA participants are significantly faster than the LAA participants. Finally, the analysis by subject revealed a main effect of verb ($F_1(1,61) = 5.612, p < .021$, partial $\eta^2 = .084$), qualified by a construction $\times$ verb interaction ($F_1(1,61) = 4.554, p < .037$, partial $\eta^2 = .069$); however, neither of these was significant in the analysis by item. This suggests that using a verb in a construction with which it is not normally associated incurs a higher cost for passive sentences than for actives. As shown in Table 4, differences in response times between $AV_A$ (active constructions containing active-attracting verbs) and $AV_P$ (active constructions containing passive-attracting verbs) conditions are negligible for both groups (in fact, response times
are faster on the AV$_P$ condition). However, paired $t$-tests reveal that both groups’ response times were significantly faster for the PV$_P$ (passive sentences containing passive-attracting verbs) condition than the PV$_A$ (passive sentences containing active-attracting verbs) condition: for HAA participants, $t(30) = 1.996$, $p = .050$, $d = 0.2$; for the LAA group, $t(31) = 2.049$; $p = .049$, $d = 0.2$.

The reaction time results show that the HAA group was faster than the LAA group overall. This is to be expected and could be due to a number of factors (e.g., greater overall processing speed, greater familiarity with computers, or simply better literacy skills). Also, as expected, both groups had slower reaction times for passives. This could be attributed to entrenchment (actives are much more frequent than passives), sentence length (passives are slightly longer than actives), morphological complexity (passives contain additional morphological markers), or, in a generative framework, to the fact that passives are believed to involve movement. The processing of passives was speeded up with verbs that are attracted by the construction, showing that participants have lexically specific knowledge about passives. As anticipated, there was no corresponding effect for actives, which is likely to be a ceiling effect: processing times for the active are so short that there is no room for other influences to manifest themselves.

4.3 Relationship between speed and accuracy

To determine whether there is a relationship between speed and accuracy, we computed Spearman correlations between these two variables for all sentences and for actives and passives separately. The relevant figures are presented in Table 5. The correlation
coefficients range from weak and statistically insignificant to moderately strong and highly significant. Crucially, however, they are all negative, indicating that the more accurate participants also responded faster than less accurate participants. Thus, our data show no evidence of a speed-accuracy trade-off.

5. Discussion
The accuracy data reported here replicate earlier research suggesting that many speakers with relatively little schooling have problems interpreting passives. The overall performance on passives in the LAA group (86% correct) is similar to that observed in Street & Dąbrowska (2010), in spite of the fact that the testing procedure (naming the ‘doer’ or ‘acted-on’) was much more demanding than the picture selection task used in that study.

Do these differences in decision accuracy reflect differences in underlying linguistic representations, or could they be attributed to linguistically-irrelevant factors such as willingness to cooperate with the experimenter, amount of experience with formal testing or ability to perform the experimental task? In our view, appeals to performance factors as an explanation of the decision accuracy results are highly unsatisfactory. First, issues surrounding ‘test-wiseness’ should be evident across all constructions; yet the decision accuracy of the LAA group was at ceiling in the active conditions. Second, the test trials were preceded by a supervised practice trial conducted to establish that participants did, in fact, understand the task. Finally, all participants were extremely co-operative.
As explained in the introductory section, the main purpose of the study was to determine whether the less educated speakers’ difficulties with the passive can be attributed to reliance on lexically specific representations or processing heuristics. Specifically, we tested two predictions derived from usage-based models (that there should be a processing advantage for passive-attracting verbs used in passive sentences and that this advantage should be stronger for the LAA group) and a further prediction derived from two-stage processing models such as LAST (that faster reaction times should be associated with lower accuracy).

Our results support the idea that speakers make use of lexically-specific representations. As we have seen, both groups responded faster to passive sentences containing passive-attracting verbs than to passive sentences with active-attracting verbs. This is consistent with earlier research suggesting that lexically specific knowledge survives into adulthood (Garnsey et al., 1997, Holmes et al., 1989, MacDonald, Perlmutter & Seidenberg, 1994, MacDonald & Seidenberg, 2006, Trueswell et al., 1993, Trueswell & Tanenhaus, 1994).

However, prediction 2 (for the HAA group, $PV_P > PV_A$; for the LAA group, $PV_P >> PV_A$) was not confirmed. There was no interaction between construction, verb type and group: although the LAA participants were somewhat slower overall than the HAA group, they did not experience a particular difficulty with active-attracting verbs in the passive. Moreover, they were equally accurate with passive sentences with passive-attracting verbs and passive sentences with active-attracting verbs (85.4% and 85.9% correct respectively). Thus, our results do not support the proposal made by Dąbrowska & Street (2006) that education-related differences in the comprehension of passive
sentences arise as a consequence of LAA participants relying on verb-specific schemas to a greater extent than more educated participants.

Finally, we found that the HAA participants were significantly faster as well as more accurate at processing passive sentences; and in both groups the faster participants were also more accurate. Thus, prediction 3, that there should be a positive correlation between accuracy and reaction time for passives, was also not confirmed. This finding is inconsistent with the proposal that individual differences in the comprehension of passive sentences are attributable to differences in parsing ability, specifically the contention that mistakes in the comprehension of passives arise when listeners abort processing after the pseudoparse and never compute the true parse. Such an account would predict a speed-accuracy trade-off; but there is no evidence of such a trade-off in our data: in fact, we found the opposite.

Thus, neither of the two explanations proposed in the introduction can account for the observed results. Our findings suggest that both groups of participants have verb-general as well as verb-specific representations and both process the sentences in the same way (although the LAA participants are somewhat slower). The difference, we suggest, lies in the degree of entrenchment of these representations, which in turn is a function of the amount of experience with the passive construction. Passive sentences are considerably more frequent in written texts (particularly formal written texts) than in speech; and since more educated participants tend to read more, their passive constructions are better entrenched, and hence accessed more reliably, which results in faster and more accurate performance. This is true for verb-specific constructions (since they have more experience with individual passive-attracting verbs) as well as verb-
general constructions (since they have experienced more verb types in the passive). Entrenchment is clearly a matter of degree: hence, performance on relatively infrequent structures such as the passive varies considerably, particularly in the LAA group, where individual scores on passive sentences ranged from 8% to 100%.

The results reported here support two fundamental claims of usage-based models: that much of our linguistic knowledge is lexically specific, and that frequency plays a crucial role in shaping speakers’ mental grammars. More experience with a particular construction results in greater entrenchment, and hence more reliable performance. Our results are also compatible with constraint-based lexicalist (CBL) models (MacDonald, Perlmutter & Seidenberg, 1994; MacDonald & Seidenberg, 2006; Trueswell & Tanenhaus, 1994). CBL models share many assumptions with usage-based models, and indeed can be regarded as a type of usage based model. Proponents of such approaches see sentence comprehension as involving rapid integration of a variety of probabilistic constraints emerging from the lexical properties of individual words, the relative frequency of the verb in different constructions, the frequency of the constructions themselves, thematic fit of the verb’s arguments, and information derived from the preceding discourse and the non-linguistic context. According to CBL models, lexical information, including grammatical preferences of individual verbs, is a particularly good cue, in that it is frequently available and more reliable than most other cues. Our results indicate that speakers know which verbs are strongly associated with the active and the passive and are able to use this information on-line; consequently, passive sentences with passive attracting verbs are processed faster than passives with active-attracting verbs.
6. Conclusion

As in previous studies (e.g., Dąbrowska & Street, 2006, Street & Dąbrowska, 2010), the low academic attainment participants performed well below ceiling (and, in some cases, at or even below chance) on a task tapping comprehension of passive sentences. Their relatively poor performance cannot be explained by task demands since both groups performed at ceiling on actives.

Both groups processed passive sentences with passive-attracting verbs faster than passives with active-attracting verbs, suggesting that both groups have lexically specific knowledge about which verbs are used in the passive. This finding adds to the substantial body of research within the constraint-based lexicalist framework suggesting that speakers are highly sensitive to the distributional characteristics of individual words. Significantly, the processing advantage for passives with passive-attracting verbs was equally strong in both groups, but there were no differences between the number of correct responses for passives with passive-attracting and passive-repelling verbs in either group. This shows that the LAA group’s relatively poor performance on passives cannot be explained by appealing to differences in the availability of a verb-general passive construction (as proposed by Dąbrowska & Street, 2006). Instead the results suggest that both groups have verb-specific and verb-general constructions. Significantly, however, the LAA group’s knowledge about the passive is less well entrenched, and hence accessed less reliably.

As pointed out in the introductory section, the claim that all first language learners acquire essentially the same grammar is one of the strongest arguments for an innate
language faculty, and one of the “facts” that the Universal Grammar hypothesis was
supposed to explain. The results discussed in this paper add to a growing body of
evidence suggesting that the convergence argument is based on a false premise: there are,
in fact, substantial differences in native speakers’ mastery of a number of grammatical
constructions, and, moreover, these differences appear to be related to differences in
linguistic experience. It does not necessarily follow from this that Universal Grammar
does not exist: one can argue for strong innate constraints on language learning on other
grounds (e.g., poverty of the stimulus). It is also possible that the individual differences
observed in this and other studies reflect differences in the innate biases for language
learning. Indeed, given that there are considerable individual differences in most
genetically determined traits, we would expect variation in the innate endowment for
language as well.

Many linguists will argue that our less-educated participants’ difficulties in
interpreting passive sentences are facts about linguistic performance rather than
competence (and hence irrelevant to the truth or falsity of the Universal Grammar
hypothesis). Whether we accept this argument depends, of course, on where we draw the
line between competence and performance: in a usage-based theory, systematic
differences in comprehension accuracy and processing speed are thought to reflect
differences in entrenchment, and hence facts about speakers’ linguistic representations,
not just their use of these representations in processing. Wherever we draw the line, it is
clear that the fact that some adults perform at chance, or even below chance, on a task
tapping comprehension of passive sentences is problematic for theories which claim that
all first language learners master the basic constructions of their language at a young age.
Achieving full mastery of the passive (and other constructions – see introduction) requires a considerable amount of experience with this construction.
Appendix I

List of sentences used in one version of the test

**AV\(_A\)** (active sentences with active-attracting verbs)

Thomas bit Roger
James grabbed Peter
Mary punched Sandra
Emma touched Jane
Robert shook John
Rachel beat Sally

**PV\(_A\)** (passive sentences with active-attracting verbs)

Sally was bitten by Rachel
Sandra was shaken by Mary
Peter was beaten by James
John was punched by Robert
Jane was grabbed by Emma
Roger was touched by Thomas
AVP (active sentences with passive-attracting verbs)

Roger fed Thomas
Peter kicked James
John flicked Robert
Sandra attacked Mary
Jane injured Emma
Sally hurt Rachel

PV_P (passive sentences with passive-attracting verbs)

Thomas was hurt by Roger
James was injured by Peter
Robert was attacked by John
Mary was flicked by Sandra
Emma was kicked by Jane
Rachel was fed by Sally
Appendix II

Written instructions

You will see sentences like this one: Bill hit Jamie.

In this sentence Bill is the “doer” - the person doing the action (he hits Jamie).
Jamie is the “acted on” - the person affected by the action (somebody hits him).

Your task is to read each sentence and then name either: the “doer” - by pressing “D” on
the keypad OR the “acted on” - by pressing “A” on the keypad.

EXAMPLE: for this sentence you would press “D” because Justin is the “doer.”

Justin scratched Steven

but for this sentence you would press “A” because Jacky is “acted on”.

Tracey hit Jacky
You have as much time as you need to answer: the sentence will remain on screen until you select either “D” or “A”.

After your selection, you will see this sign +++ for 2 seconds. Then the next sentence will appear.

The first 4 sentences are to practice.

Press SPACEBAR to begin Practice Session.
Acknowledgements

This research was partially supported by the AHRC (award no. 06/125939 to the first author).
References


Linguistics 35, 735–766.


Language Sciences 28, 604–615.


Tables

Table 1: Relative frequencies in the active and passive, overall frequencies and p-value of attraction to construction of the verbs used in the experiment

<table>
<thead>
<tr>
<th>Verb</th>
<th>Frequency in active (ICE-GB)</th>
<th>Frequency in passive (ICE-GB)</th>
<th>Frequency in active (BNC)</th>
<th>Frequency in passive (BNC)</th>
<th>Overall frequency (BNC)</th>
<th>Fisher-Yates p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type A: Active-attracting</strong></td>
<td>Attraction to Actives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Touch</td>
<td>30</td>
<td>3</td>
<td>90</td>
<td>10</td>
<td>10836</td>
<td>0.121</td>
</tr>
<tr>
<td>Beat</td>
<td>38</td>
<td>8</td>
<td>60</td>
<td>30</td>
<td>9440</td>
<td>0.526</td>
</tr>
<tr>
<td>Shake</td>
<td>28</td>
<td>2</td>
<td>98</td>
<td>2</td>
<td>8903</td>
<td>0.068</td>
</tr>
<tr>
<td>Bite</td>
<td>6</td>
<td>1</td>
<td>76</td>
<td>24</td>
<td>3414</td>
<td>0.623</td>
</tr>
<tr>
<td>Punch</td>
<td>4</td>
<td>0</td>
<td>74</td>
<td>26</td>
<td>2319</td>
<td>0.445</td>
</tr>
<tr>
<td>Grab</td>
<td>7</td>
<td>0</td>
<td>88</td>
<td>12</td>
<td>2936</td>
<td>0.242</td>
</tr>
<tr>
<td>Mean frequency</td>
<td>18.8</td>
<td>2.3</td>
<td>81.0</td>
<td>17.3</td>
<td>6308</td>
<td></td>
</tr>
</tbody>
</table>

| **Type P: Passive-attracting** | Attraction to Passives | | | | | |
| Kick   | 6                            | 3                            | 55                        | 45                         | 5167                    | 0.219                |
| Feed   | 26                           | 17                           | 40                        | 60                         | 7479                    | 0.001                |
| Injure | 4                            | 19                           | 17                        | 83                         | 2508                    | <0.001               |
| Attack | 35                           | 21                           | 45                        | 55                         | 16411                   | 0.001                |
| Flick  | 5                            | 5                            | 70                        | 30                         | 1254                    | 0.023                |
| Hurt   | 15                           | 8                            | 40                        | 50                         | 5239                    | 0.046                |
| Mean frequency | 15.2              | 12.2                        | 44.5                      | 53.8                      | 6343                    |                      |
Table 2: *Examples of test sentences*

<table>
<thead>
<tr>
<th>Verb</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active</strong></td>
<td><strong>Passive</strong></td>
</tr>
<tr>
<td>$V_A$ (Active-attracting)</td>
<td><em>James grabbed Peter</em></td>
</tr>
<tr>
<td>$V_P$ (Passive-attracting)</td>
<td><em>Jane injured Emma</em></td>
</tr>
</tbody>
</table>
Table 3: **Proportion of correct responses for each condition by group**

<table>
<thead>
<tr>
<th>Group</th>
<th>( AV_A ) (Active-attracting verb in active sentence)</th>
<th>( AV_P ) (Passive-attracting verb in active sentence)</th>
<th>( PV_A ) (Active-attracting verb in passive sentence)</th>
<th>( PV_P ) (Passive-attracting verb in passive sentence)</th>
<th>Active (collapsed ( AV_A ) and ( AV_P ))</th>
<th>Passive (collapsed ( PV_A ) and ( PV_P ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAA (N=31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>99 (3)</td>
<td>98 (6)</td>
<td>95 (8)</td>
<td>98 (6)</td>
<td>99 (3)</td>
<td>98 (6)</td>
</tr>
<tr>
<td>Median (Range)</td>
<td>100 (83-100)</td>
<td>100 (83-100)</td>
<td>100 (83-100)</td>
<td>100 (83-100)</td>
<td>100 (92-100)</td>
<td>100 (83-100)</td>
</tr>
<tr>
<td>LAA (N=32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>98 (7)</td>
<td>98 (7)</td>
<td>86 (20)</td>
<td>85 (20)</td>
<td>98 (6)</td>
<td>86 (18)</td>
</tr>
<tr>
<td>Median (Range)</td>
<td>100 (67-100)</td>
<td>100 (67-100)</td>
<td>91 (0-100)</td>
<td>100 (17-100)</td>
<td>100 (67-100)</td>
<td>91 (9-100)</td>
</tr>
</tbody>
</table>

Table 4. **Mean response times and SDs (in milliseconds) of sentences by group**

<table>
<thead>
<tr>
<th>Condition</th>
<th>High academic attainment (N=31)</th>
<th>Low academic attainment (N=32)</th>
<th>Overall Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>( AV_A ) (Active-attracting verb in active sentence)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3495 (863)</td>
<td>4170 (1225)</td>
<td>3833</td>
</tr>
<tr>
<td>( AV_P ) (Passive-attracting verb in active sentence)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3484 (982)</td>
<td>4156 (1316)</td>
<td>3820</td>
</tr>
<tr>
<td>( PV_A ) (Active-attracting verb in passive sentence)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>4720 (1492)</td>
<td>5688 (1496)</td>
<td>5204</td>
</tr>
<tr>
<td>( PV_P ) (Passive-attracting verb in passive sentence)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>4436 (1396)</td>
<td>5385 (1415)</td>
<td>4910</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>4034</td>
<td>4850</td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Correlations between decision accuracy and response time

<table>
<thead>
<tr>
<th>Condition</th>
<th>High academic attainment (N=31)</th>
<th>Low academic attainment (N=32)</th>
<th>All participants (N=63)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actives</td>
<td>-.28</td>
<td>-.28</td>
<td>-.38**</td>
</tr>
<tr>
<td>Passives</td>
<td>-.09</td>
<td>-.41*</td>
<td>-.42***</td>
</tr>
<tr>
<td>All sentences</td>
<td>-.16</td>
<td>-.30</td>
<td>-.40***</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level

** Correlation is significant at the 0.01 level

*** Correlation is significant at the 0.001 level
Notes

1 Passives are 4-5 times more frequent in writing than in speech (Roland et al. 2007). The difference in frequency is even more pronounced for full passives (i.e., passives with an agentive adjunct). In the British National Corpus, for example, the average frequency of the full passive is 63 per million words for written texts and about 9 per million in speech – a sevenfold difference in frequency.

2 The NVN strategy refers to the comprehender’s tendency to assume that the subject of a sentence is also the agent of a particular action and the object of the same action is theme. It is argued that this occurs because the majority of sentences in English conform to this pattern. Consequently, those sentences that require the patient thematic role to be assigned before the agent thematic role (e.g., passives, object-clefts) prove more difficult to process than those that conform to the more common NVN pattern (see Ferriera 2003, p. 25).

3 These figures were obtained by averaging across several different conditions in two experiments.

4 The number of items per condition (6) was kept relatively small to keep the task as short as possible in order to ensure that the low-education participants remained engaged; it is, however, in line with previous studies using a similar methodology, e.g. Ferreira (2003).

5 The Fisher-Yates p-values and frequencies in the ICE-GB corpus are taken from unpublished data collected by Stefan Gries & Anatol Stefanowitsch. We thank these researchers for sharing this information with us.
To ensure that the results reported here are not unduly influenced by outliers, we conducted a second ANOVA in which all observations which were more than 1.5 SDs from the mean were removed. The results were very similar to those reported in the text.