

Lexical chunking effects in syntactic processing

ARNE ZESCHEL*

Abstract

Research on syntactic ambiguity resolution in language comprehension has shown that subjects' processing decisions are influenced by a variety of heterogeneous factors such as e.g. syntactic complexity, semantic fit and the discourse frequency of the competing structures. The present paper investigates a further potentially relevant factor in such processes: effects of syntagmatic lexical chunking (or matching to a complex memorized prefab) whose occurrence would be predicted from usage-based assumptions about linguistic categorisation. Focusing on the widely studied so-called DO/SC-ambiguity in which a post-verbal NP is syntactically ambiguous between a direct object and the subject of an embedded clause, potentially biasing collocational chunks of the relevant type are identified in a number of corpus-linguistic pretests and then investigated in a self-paced reading experiment. The results show a significant increase in processing difficulty from a collocationally neutral over a lexically biasing to a strongly biasing condition. This suggests that syntagmatically complex and partially schematic templates of the kind envisioned in usage-based Construction Grammar may impinge on speakers' online processing decisions during sentence comprehension.

Keywords: Sentence processing, prefabs, usage-based model

1. Introduction

Research on language processing has shown that comprehenders have temporary difficulties with the interpretation of locally ambiguous sentences like those in (1):

- (1) a. The criminal confessed his sins harmed too many people.
(Rayner and Frazier 1987)
- b. The thief searched by the police had the missing weapon.
(Trueswell 1996)
- c. The complex houses single and married students and their families. (Jurafsky 1996)

Whereas it is widely assumed that such difficulties provide valuable cues as to how the underlying processing system is organised, there is as yet no general consensus about the factors that account for the observed difficulties (cf. Tanenhaus and Trueswell 1995 for an overview). Most consonant with usage-based approaches to language are so-called constraint-based models of the ambiguity resolution process which argue for an immediate interaction and rapid integration of different information sources: in contrast to syntax-centered two-stage models, these approaches are non-modular and accord central importance to matters of usage frequency and psychological entrenchment, both assumptions that are well in keeping with central tenets of cognitively oriented versions of Construction Grammar (Bybee 2006; Goldberg 2006; Langacker 2000).

The present paper adds to this research by evaluating the role of a factor that has received comparably little attention in the literature so far, viz. potential effects of collocational chunking on the ambiguity resolution process. Focusing on the so-called “DO/SC” (sometimes also called “NP/S”) ambiguity in which a post-verbal NP is temporarily ambiguous between a direct object and the subject of an embedded clause (cf. 1a), initial evidence from a self-paced reading experiment is presented which suggests that complex and partially schematic templates of the kind envisioned in usage-based Construction Grammar impinge on subjects’ syntactic processing decisions.

2. A usage-based perspective on sentence processing

One of the foundational assumptions of usage-based approaches is that linguistic knowledge is heavily redundant, with abstract schemas coexisting with specific instances of the relevant type that have independent unit status themselves (provided they are sufficiently entrenched). From a processing perspective, this raises the following question: if there are several elements in the construction that could be invoked as the categorising structure for a given target expression, then which of these will actually be

chosen? The question is important since different candidates may warrant different predictions as to how the utterance will unfold further, meaning that the choice of a particular candidate structure at the expense of others may have consequences for later processing decisions.

As indicated in the introduction, traditional psycholinguistic approaches to sentence processing can be broadly distinguished into two types of models. Prototypical instances of the first type are serial and modular: such models assume that parsing decisions are initially guided by considerations of syntactic complexity alone, and that attachment ambiguities are resolved through general heuristics such as “minimal attachment” and “late closure” without recourse to non-syntactic information unless the initial analysis fails, in which case the parser has to backtrack and reanalyse (Ferreira and Clifton 1986; Frazier 1987; Frazier and Fodor 1978). The second class of models is typically parallel and interactive: here, comprehenders are assumed to employ constraints from a variety of different sources from the outset, with several different analyses competing for selection (e.g., MacDonald et al. 1994; Trueswell et al. 1993; Trueswell et al. 1994).

Langacker (2000) sketches a usage-based perspective on the selection of linguistic categorising structures that is compatible with models of the latter type. Specifically, Langacker assumes that only a single categorising structure will be selected in the end, and that the choice of this element is influenced by three types of factors: entrenchment (i.e. relative degree of resting activation and routinization of the competing alternatives), contextual priming (plausibility within the present discourse context) and overlap (degree of structural similarity with the specific target at hand). While the first two of these three factors have been extensively studied in the ambiguity resolution literature (cf. e.g. Cuetos, Mitchell and Corley 1996; Jurafsky 1996; Pickering, Traxler and Crocker 2000; Trueswell 1996 on aspects of frequency and entrenchment; Garnsey et al. 1997; Hare et al. 2003, 2004; Tanenhaus et al. 2000; Trueswell et al. 1994; Wiechmann, this volume on semantic fit and contextual plausibility), the third factor, “overlap”, has received considerably less attention.

The present paper addresses this factor, with “overlap” understood as similarity of the target to a larger composite structure that is hypothesised to have psychological unit status and which could thus be invoked as a single categorising structure holistically. A

possible explanation for the relative neglect of this question in previous research is that issues of syntax and sentence processing are not commonly thought about in terms of prefabricated formulae because they are seen as the provenance of free and unrestricted combinatoriality in language, with highly general syntactic rules being the most salient manifestation of what Sinclair (1987) has called the “open-choice principle” in language. As a consequence, sentence processing is usually assumed to involve an incremental build-up from atomic units rather than an amalgamation of more or less complex structures that may already be retrieved en bloc. As indicated above, however, usage-based approaches assume that speakers do indeed memorize such internally complex chunks (regardless whether they are predictable or not), and that the generalisations that speakers extract from structurally similar elements in their repertoire (e.g. VP → V NP as a generalisation over e.g. *ask a question, hit the post, pull strings* etc.) are in fact epiphenomenal (in the sense that they are merely implicit in a set of stored exemplars, cf. Langacker 2000). Prefabricated chunks of various grain sizes thus play an important role in the model, and it seems reasonable to hypothesise from here that they are also relevant for processing.

So far, there has been little experimental work on this issue, even though it has not remained unnoticed: for instance, Elman et al. (2004) mention the possibility that formulaic sequences such as *Mr. and Mr. Smith proudly announce...* may induce processing biases that differ from that of the verb as viewed “in isolation” (here: a DO-preference in spite of the overall SC-preference of *announce*). Apart from such potentially biasing effects of *prior* context, i.e. preverbal material, one would also expect the attachment of the ambiguous noun phrase to be sensitive to the concrete lexical identity of its head: specifically, encountering an NP that is a direct object collocation of the respective verb should privilege a DO-analysis of the developing structure, even if the verb in isolation otherwise favours SC. Moreover, one would expect that any additional syntagmatic cues for this reading should increase the hypothesised effect.

Arguably, aspects of entrenchment, contextual plausibility and overlap may be difficult to disentangle in practice: what is frequent in the input is usually frequent for a good semantic reason, and individual frequent combinations are of course likely to be stored. Hence, it can be expected that collocating nouns in VN-sequences occur in this position more frequently than expected, that they will

allow a semantically coherent interpretation *and* that the entire sequence may in fact be stored as a prefab. Nevertheless, habitual co-occurrence is still not the same as semantic plausibility. Specifically, one can expect routinized co-occurrences to be semantically plausible for some reason or other, but not necessarily vice versa (i.e. there are all sorts of things that can be plausibly confessed, but not all of the corresponding nouns are habitual collocates of the verb *confess*). This is where the present study comes in: the experiment reported in section 3 attempts to tease apart effects of contextual plausibility on the one hand and overlap on the other hand by comparing the processing of putative prefabs to the processing of presumably non-stored VN-sequences that contain a close semantic variant of the collocating noun.

To my knowledge, there is no previous research on the influence of collocation effects on syntactic processing in the sense outlined above. On a more general level, however, there is experimental evidence that comprehenders do indeed use latent statistical cues from the linguistic context to speed up comprehension. For instance, McDonald and Shillcock (2003) present eyetracking evidence that transitional probabilities between verbs and nouns affect gaze duration in reading, concluding that “the brain is able to draw upon statistical information in order to rapidly estimate the lexical probabilities of upcoming words: a computationally inexpensive mechanism that may underlie proficient reading” (McDonald and Shillcock 2002: 648). The present study combines corpus-linguistic and experimental methods to investigate whether such information also influences syntactic processing, and how possible chunking effects of this type relate to item-based preferences pertaining to the verb when viewed in isolation.

3. Prefabs in sentence comprehension

Effects of syntagmatic lexical chunking on syntactic processing were investigated in a self-paced reading experiment. Subjects read different types of locally ambiguous sentences that ultimately turned out to involve sentential complementation. Target items were of three types:

- stimuli that could not be said to privilege a DO-analysis due to lexical chunking effects because they did not involve a collocating VN-pair at all (even though the

respective noun was a near-synonym of the collocating noun and hence semantically plausible)

- stimuli that supported a transitive DO-analysis before the disambiguation region by involving a DO-collocating noun, yet no additional pointers to the ultimately wrong DO-analysis (i.e. the “core” of the hypothesised DO-prefab alone)
- stimuli that strongly supported a transitive DO-analysis before the disambiguation region by involving a DO-collocating noun and additional preverbal cues for the collocating DO-chunk (i.e. a hypothesised complex prefab)

Suitable stimuli were constructed on the basis of a number of corpus-linguistic pretests. These pretests departed from a list of 16 SC-biased verbs that were taken from an earlier study (Garnsey et al. 1997) and explored potentially interesting DO-uses of these items in the British National Corpus (BNC).¹ In the first step, each verb was concordanced in all relevant forms, and frequency counts for nouns occurring at positions R-1 and R-2 were summed. For each verb, the list of co-occurring nouns was then sorted for frequency and the top three VN-pairs were concordanced anew, this time with a larger span of up to five words in between verb and noun in order to also capture instances involving e.g. disjuncts and modification. Finally, three VN-collocations from the resulting concordances were chosen that looked promising for present concerns.² The target items of the present study thus selected were the three combinations *admit—defeat*, *believe—luck*, and *prove—worth*.

It was ensured that the non-collocating combinations in the first condition were nevertheless attested in naturally occurring English text (if not in the BNC, then at least on .uk sites on the web). Examples of such near-synonymous combinations are given in (2):

- (2) a. Davids proved his value for the Ajax team again.
http://gov-certificates.co.uk/birth/certificate/Edgar_Davids
 (last accessed 25-1-2008)
- b. Tories can't ever admit losing without stamping their feet up and down and howling and bawling ‘not fair’, ‘not fair’.
<http://chat.thisislondon.co.uk/london/threadnonInd.jsp?forum=18&thread=220080>
 (last accessed 25-1-2008)
- c. Darren Moore could scarcely believe his fortune when he

headed gently in amid a motionless Hull defence.
http://football.guardian.co.uk/Match_Report/0,,72111,00.html
 (last accessed 25-1-2008)

3.1. *Corpus-linguistic pretests*

3.1.1. *Methods*

Association strength computations for the presumed collocations were calculated in the form of a covarying collexeme analysis (Stefanowitsch and Gries 2005) in order to apply a measure that is sensitive to syntactic structure. In other words, only those occurrences of e.g. *admit* followed by *defeat* (within a certain preset span) were counted as a relevant hit in which the noun was actually the direct object of the verb. This was necessary because the very existence of the ambiguity illustrates that mere (near-) adjacency of two words in a sequence does not say anything about the structural relations between these items. For instance, in the following hits for queries of the type [V]...w5...[N], the supposedly DO-collocating noun is not a direct object of the verb:

- (3) a. ...Dowens admitted the defeat left him a little flat...
 [BNC A9U]
 b. Nobody but a fool who believes his luck lies around the corner could... [BNC ART]
 c. Pieces which have proved to be of enduring worth have passed... [BNC FPY]

Covarying collexeme analyses permit the identification of significant associations between words in different slots of one and the same grammatical construction. Unfortunately, the method requires either a parsed corpus or extensive manual post-editing of the data. Since the BNC is not syntactically annotated and balanced parsed corpora such as ICE-GB are much too small to investigate the comparably rare bigrams that are at issue here, samples had to be drawn for the figures that could not be exhaustively coded by hand.³ Specifically, these were the frequencies of

- the target verb co-occurring with all other nouns in the transitive construction
- the target noun co-occurring with all other verbs in the transitive construction

- the transitive construction in the corpus at large

Table 1 illustrates the actual calculation of these values on the example of *admit—defeat*. The plain format figures in the table were obtained directly from the corpus/corpus samples, the italicized ones were arrived at by subtraction (see below):

Table 1 *Input figures for admit—defeat*

	Noun N in the transitive Cxn	All other nouns (+transitive)	Totals
Verb V in the transitive Cxn	<i>[admit defeat]_{trans}</i> 58	<i>[admit N_{-defeat}]_{trans}</i> 2065	2123
All other verbs (+transitive)	<i>[V_{-admit} defeat]_{trans}</i> 716	<i>[V_{-admit} N_{-defeat}]_{trans}</i> 3,744,787	3,745,503
Totals	774	3,746,852	3,747,626

To begin with, the frequency of *[admit defeat]_{transitive}* in the upper left cell (58) was obtained by syntactically analysing the 62 raw hits of the corpus query. The frequency of *[V_{-admit} defeat]_{transitive}* in the lower left cell was obtained by analysing a sample of 346 examples out of the 3,476 total occurrences of *defeat* that are tagged as a noun in the BNC.⁴ Specifically, I first identified the number of hits in this sample in which *defeat* functioned as the direct object in a transitive construction (77); second, the proportion of transitive direct object uses in the sample was extrapolated to the overall population of nominal *defeat* (774); third, the number of hits for *[admit defeat]_{transitive}* was subtracted from this figure, thus giving the estimated number of transitive constructions consisting of a verb other than *admit* and *defeat* as the direct object (716). The same procedure was applied in order to arrive at the estimated number of tokens for *[admit N_{-defeat}]_{transitive}* (2065). The figure in the lower right cell (all transitive constructions in the BNC which have neither *admit* in the V-slot nor *defeat* in the N-slot) was arrived at in two steps: first, the figures for *[admit N]_{transitive}* and *[V defeat]_{transitive}* (i.e. the known row and column totals) were subtracted from the total number of transitive constructions in the BNC, thus giving the last missing row- and column totals.⁵ Once these were in place, the figures for *[admit N_{-defeat}]_{transitive}* and *[V_{-admit} defeat]_{transitive}* could be subtracted from these results, thereby giving the value in the final missing

cell. On the basis of the completed table, it was then possible to calculate the expected frequency of [*admit defeat*]_{transitive} in the BNC (0.4) and to evaluate the difference between the observed and the expected value.

Once association strengths were calculated in this manner, all attestations of the three target items in the BNC were subject to a detailed manual coding of their syntagmatic context profile within a span of \pm five words (with the verb as node). Full manual post-editing was applied for two reasons: first, the POS-tagging of the BNC is not 100 percent reliable, so that the co-occurrence figures obtained from automatically generated collocate lists are not necessarily correct.⁶ Second, even if tagging were 100 percent reliable, automatically generated collocate lists would still remain an imperfect approximation of the grammatical co-occurrence properties of the investigated items because they do not take syntactic structure into account. For instance, *admit defeat* is often found with adverbs such as *finally* or *never* that typically (4.a, b)—though not always (4.c, d)—occur immediately before the verb:

- (4) a. ...the man who believed this would never admit defeat...
[BNC CAW]
b. ...she finally admitted defeat when... [BNC BP4]
c. Some people never do admit defeat. [BNC G3D]
d. ...but was finally having to admit defeat and... [BNC G3D]

It would of course be desirable to quantify co-occurrences with such elements in the same way in which the association strength between the verb and the noun was quantified, i.e. using standard collocation methods by assessing the association of each item in the string with the constructional slot in question. On the other hand, these methods are only applicable to the constitutive slots of a given construction, which makes it difficult to accommodate optional elements such as e.g. negators or adverbial adjuncts. As a result, behavioural profiles were identified in a more informal way that relied on raw frequency of co-occurrence instead (giving observations of the type “X percent of the instances of *admit defeat* involve negation”, “X percent involve an aspectual adverb such as *finally*” etc.). Combining different such observations, lexico-grammatical context profiles were identified through detailed manual annotation of 58 relevant (i.e. transitive direct object) hits for *admit—defeat*, 77 observations of *believe—luck*, and likewise 77 instances of *prove—worth*.

3.1.2. Results

The results of the association strength computations are reported in Table 2 (values indicate probability of error that the association is non-chance as computed by the Fisher-Yates exact test, cf. Stefanowitsch and Gries 2003, 2005 for discussion):

Table 2 Association strengths of the three VN-bigrams

Bigram	<i>admit – defeat</i>	<i>believe – luck</i>	<i>prove – worth</i>
<i>p</i>	2.51e-101***	5.69e-125***	4.12e-171***

Regarding the syntagmatic periphery of the hypothesised chunks, the following templates (comparable to the “compound lexical items” proposed in Sinclair 1996) emerged from the corpus pretests:

- (5) a. *admit—defeat*
 NP + (ADV) + ({OBLIGATION}) + ADMIT + *defeat*
- b. *believe—luck*
 NP_{*i*} + {ABILITY} + NEG + BELIEVE + POSS_{*i*} +
 (*good/bad*) + *luck*
- c. *prove—worth*
 ({POSS./TRANSFER}) + NP_{*i*} + (DET *chance to*) +
 PROVE + POSS_{*i*} + *worth*

As for the notation, elements in brackets are optional and elements in curly brackets represent semantic categories with variable lexical encoding. Hence, (9.a) indicates that transitive DO-uses of *admit defeat* consist of a subject NP that is typically followed by a particular kind of adverb (usually an aspectual one like *finally*, *eventually* or *never*, but there are also some manner items like *reluctantly*), followed by an element signalling OBLIGATION (typically *have to*, but also *must*, *be forced to* etc.), followed by a form of *admit*, followed by the direct object noun *defeat*. In the case of *believe—luck*, the subject of *believe* is commonly followed by an element signalling ABILITY (typically *can*), followed by a negative element such as *not*, *hardly* or *scarcely*, followed by a form of *believe*, followed by a possessive pronoun that is coreferential with the subject, followed by the direct object *luck*. The dominant usage pattern for transitive *prove—worth* is a little more complex since the bigram is typically embedded in an infinitival construction in which the referent of an NP is said to

have (or be given) a *chance* (*opportunity* etc.) to prove his or her worth in a particular respect (or to the benefit of a certain third party). Some representative examples of each pattern are given in (6)-(8):

- (6) a. ...the man who believed this would never admit defeat.
[BNC CAW]
b. Despite soldiering on for three days she finally admitted defeat when... [BNC BP4]
c. ...but was finally having to admit defeat and accept powerlessness. [BNC G3D]
- (7) a. Juliet couldn't believe her luck. [BNC JY0]
b. He could hardly believe his bad luck though... [BNC CEP]
c. ...she fitted so exactly that Wycliffe could scarcely believe his luck. [BNC GWB]
- (8) a. ...and we are giving him the chance to prove his worth as a footballer. [BNC K5A]
b. ...he will not feel the need to prove his worth, either to himself or... [BNC GVF]
c. ...had an extended early opportunity to prove his worth as... [BNC BN9]

Using the templates in (5), it was now possible to construct appropriate stimuli for the following reading experiment.

3.2. *Reading experiment*

3.2.1. *Subjects*

35 participants took part in the experiment. Subjects were students at the Universities of Bremen, Sheffield, Salford and Michigan.⁷ All participants were native speakers of English. Further demographic characteristics such as sex and age were not recorded since they were not deemed relevant for the purpose at hand. Subjects were not paid for participation in the experiment.

3.2.2. *Materials*

The experiment was conducted on standard personal computers with Microsoft Windows XP operating systems. Stimuli were presented using "Self Paced Reading Projector" from experimentalSuite 0.9 beta for Windows, a custom-made stand-alone application programmed with Macromedia Director.

Subjects were seated at the screens and presented with eight short texts about the Football World Cup 2006 followed by a timed yes-no comprehension question. Texts 4, 6 and 8 contained the actual test items in different orders, with the remaining texts serving as distractors. In the target items, critical passages were preceded by one to three sentences establishing an appropriate discourse context, followed by the sentence containing the critical passage, followed by one or two additional sentences before the comprehension question. Texts were designed to be as ecologically valid as possible by assembling them (as far as possible) from pieces of real-life British sports reporting. As an illustration, the set of experimental stimuli for *prove—worth* is reproduced below:

After the first two matches, Brazilian superstar Ronaldo was criticised a lot for being overweight, slow and lacking determination. Coach Carlos Alberto Parreira stayed stubbornly loyal, though, promising his centre forward a place in the starting line-up for the next match.

- a. “Ronaldo will prove his value for the team...
- b. “Ronaldo will prove his worth for the team...
- c. “Ronaldo will get a chance to prove his worth for the team...

has been downplayed by the media. He is an exceptional player, and I am very confident that he will score today”. Ronaldo repaid Parreira with two goals against Japan that led his team to the knockout stages and equalled Gerd Müller’s all-time record of 14 world cup goals.

The stimulus sets used for the other two verbs are included in the Appendix. Subjects were given printed instructions which read as follows:

You are taking part in an experiment on text comprehension. You will read a number of short texts, each followed by a short statement relating to their content that you are asked to qualify as either right or wrong.

The texts are presented on a word by word basis. The current word is presented in the middle of the screen. Pressing SPACE

will advance the presentation to the next word, replacing its predecessor, until the end of the text is reached.

You can determine the pace of the presentation yourself. It is important that you pay attention to details, so simply take as much time as you need for a careful reading of the texts. Crucially, it is *not* possible to go back to earlier words.

At the end of each sentence, the string “+++” is displayed. When you are ready for the next sentence, press SPACE to move on. When you have reached the end of a text, the string “???” is displayed in the middle of the screen.

When you get to the “???” prompt, please place one finger each on the keys “cursor left” and “cursor right”. Pressing either key will prompt the test statement to appear in the middle of the screen. Please indicate whether the statement matches the contents of the preceding text *as quickly as possible* by either pressing “cursor right” (YES) or “cursor left” (NO).

After your response, the procedure is repeated for the next text until the end of the experiment is reached.

The experiment is not a quiz. All the information that you need for a correct response is supplied in the texts.

The first text is for training. Please pause after you have responded to the first test statement. The instructor will ask you if there is still anything unclear about the procedure before the experiment begins.

3.2.3. Procedure

Words were presented one at a time in 48 point yellow font in the middle of a blue screen. Subjects advanced the presentation by pressing the space bar. Comprehension questions were answered by pressing either the key “cursor left” (“no”) or “cursor right” (“yes”).

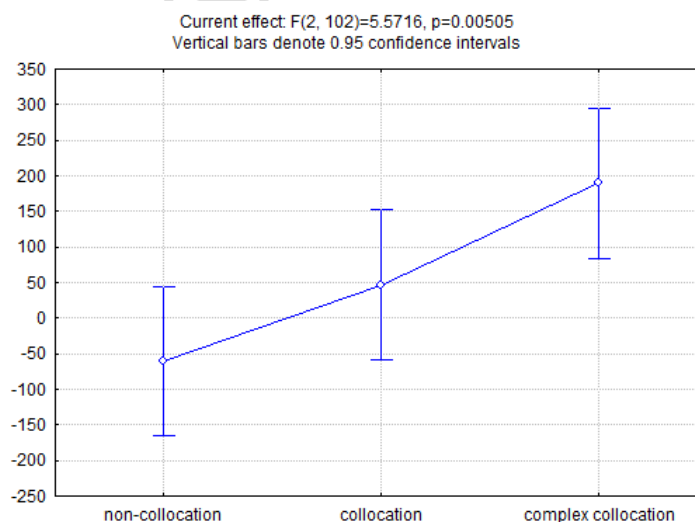
Subjects were told that the first text was for training. Once subjects had responded to the first comprehension question, the experimenter ensured that the overall procedure was clear to them and left them to complete the rest of the experiment unsupervised. Each subject saw each of the three investigated verbs only once, in one of the three conditions outlined above.

3.2.4. Results

Reading times were compared at the second word of the disambiguation region (e.g. *been* in the stimulus set reproduced in 3.2.2) in order to compensate for spillover effects. Following

suggestions by Ferreira and Clifton (1986), length-adjusted residual reading times were computed for each subject and then related to degree of collocativity in an analysis of variance. Residual reading times were obtained by computing linear regression analyses for each subject's reading performance on words of different lengths (in number of characters including punctuation) with number of characters as the explanatory variable and reading time as the dependent variable. Data from the training run and all sentence-final words were excluded. Reaction times faster than 100ms and slower than 2500ms were treated as missing data (0.93 percent). A single-group *t*-test on the regression coefficients for all subjects revealed that both the intercept and the linear component were significantly different from zero ($p < 0.001^{***}$) (cf. Lorch and Myers 1990). After computing the regressions, residuals were obtained by subtracting the predicted reading time for a given word from subjects' actual reading time for this word. Hence, positive residuals indicate slower and negative residuals faster processing of a given word than predicted by the regression. The residuals for all critical words in all three conditions (all VN-sequences were analysed together) were then submitted to an analysis of variance that indicated a significant effect of collocation ($F_{(2,102)} = 5.5716$, $p < 0.01^{**}$). Figure 1 presents the results in graphical form:

Figure 1: ANOVA results



3.3. Discussion

As predicted, Figure 1 shows a uniform increase in mean residual reading time from the syntagmatically non-biasing over the biasing to the strongly DO-biasing condition. Somewhat unexpectedly, mean residual reading times for critical words in the non-biasing condition were slightly shorter than reading times for other words of the same length. Even though it is difficult to explain why words in a potentially garden-pathing position were processed slightly *faster* than other experimental words of the same length (small as the difference may be), this result is in keeping with earlier findings that the three investigated verbs are indeed biased towards sentential complementation, i.e. there is clearly no indication of a processing difficulty at the critical position in this condition. The mean increase in processing difficulty in the collocating condition is likewise only slight, but nevertheless suggests that the presence of the collocating noun alone already works against the isolated verb bias towards sentential complementation. As expected, the most marked deviations from predicted reading times are found in the complex prefab condition with a mean increase of almost 200ms. These results can be taken as an indication that speakers do not merely memorize particular collocations of a given verb (which is uncontroversial), but that these units are at least in some cases more profitably viewed as syntagmatically complex chunks rather than as simple bigrams, and that such larger prefabs may also influence on-line syntactic processing decisions during comprehension.

Nevertheless, the present results are but a first indication that needs to be interpreted with caution. To begin with, the study did not contain an unambiguous baseline condition with an overt complementiser (e.g. *Ronaldo will prove that his worth for the team has been downplayed by the media*) as is usually included in studies of syntactic ambiguity resolution. That way, verbs could be investigated in three collocationally different conditions that could be directly compared without markedly boosting the number of experimental subjects. Irrespective of the collocation/chunking issue, however, it is well documented that the basic ambiguity effect is more pronounced for some verbs than for others, which introduces a potentially confounding factor that should be controlled for in possible follow-ups. Likewise, questions remain as to the precise definition of the complex prefabs in condition 3 and the extent to which they can be directly compared across verbs. As indicated in section 3.1.1, it is at present unclear how the

collostructional methodology developed for clearly delimited constructions such as [V NP] could be extended to larger idiom chunks with fuzzy boundaries of the type in (5) whose formal specifications gradually shade off into mere semantic preferences. Finally, the overall operationalisation of degree of syntagmatic attraction in terms of discrete levels such as “non-collocation” vs. “collocation” vs. “complex prefab” is certainly an imperfect approximation of what in reality is clearly a continuum.

All in all, however, the significant increase in mean residual reading time from condition 1 to condition 3 is a promising indication that future research in this direction may be worthwhile. Moreover, all three verbs show a uniform increase in reading time between the non-collocating and the collocating condition which suggests that the observed effect is not due to “contextual priming” in the sense of section 2 alone: since the non-collocating nouns in condition 1 are a consistently weaker cue for the DO-analysis than their synonymous variants in the collocating condition, the increase in reading time is obviously not due to semantic factors. Likewise, the difference in mean residual reading time between the collocating and the complex prefab condition cannot be accounted for by appealing to increased semantic plausibility of the DO-analysis in the latter case. Instead, the results support Langacker's (2000) assumption that “overlap” between stretches of the input and complex preassembled categorising structures is a relevant processing factor in its own right.

4. Implications

The results of the present study suggest that speakers retain memory for a variety of syntagmatic context features associated with the different usage patterns of a given verb, and that accumulating syntagmatic evidence for patterns of this type may override otherwise dominant parsing biases attaching to the verb “in isolation”. These findings are consistent with the usage-based hypothesis that

“lower-level schemas, i.e. structures with greater specificity, have a built-in advantage in the competition with respect to higher-level schemas. Other things being equal, the finer-grained detail of a low-level schema affords it a larger number of features potentially shared by the target” (Langacker 2000: 16).

On the procedural level, it seems plausible to assume that this is indeed a relevant factor that influences “pattern capture”, i.e. the question which candidate out of the initial activation set the system will actually settle to in the end. Functionally, a bias towards concreteness also has clear advantages: since speakers/hearers store whatever is sufficiently frequently encountered and hence both communicatively and cognitively routinized, accumulating evidence for a particular chunk of this type means that there is a good chance that the corresponding analysis will prove the correct guess again and thus serves to relieve (or rather bypass) further processing load probabilistically.

Coming back to the question raised in the beginning (“if there are several elements in the construction that could be invoked as the categorising structure for a given expression, then which of these candidates will actually be chosen?”), the following answer would be consistent both with the general bottom-up orientation of usage-based models and the empirical results presented above: all else being equal, hearers/readers will choose the most concrete potential categorising structure that is consistent with the currently identified input, to the effect that (more) abstract schemas will only be invoked as a kind of last resort where a more concrete standard of comparison is not available. It remains for future research to show whether this more specific hypothesis can be corroborated, and how the different factors that were found to influence syntactic ambiguity resolution should be weighted. For the moment, suffice it to say that lexical chunking effects of the type investigated in this study do seem to be one of these factors, a result that is well in keeping with usage-based assumptions about the kinds of linguistic representations that speakers store and retrieve in processing.

Received 29 November 2007

Universität Bremen

Revision received 25 January 2008

Appendix

Stimulus Set A: prove—worth

After the first two matches, Brazilian superstar Ronaldo was criticised a lot for being overweight, slow and lacking determination. Coach Carlos Alberto Parreira stayed stubbornly loyal, though, promising his centre forward a place in the starting line-up for the next match.

- a. “Ronaldo will prove his value for the team...
- b. “Ronaldo will prove his worth for the team...
- c. “Ronaldo will get a chance to prove his worth for the team...

has been downplayed by the media. He is an exceptional player, and I am very confident that he will score today". Ronaldo repaid Parreira with two goals against Japan that led his team to the knockout stages and equalled Gerd Müller's all-time record of 14 world cup goals.

Stimulus Set B: admit—defeat

England did not live up to the high expectations at home, and for most of the time, coach Sven Göran Eriksson seemed obstinate in his decision to ignore what was happening right before his eyes. In an interview after the disastrous penalty shoot-out against Portugal, Eriksson still continued to act as if he could scarcely believe that his time was up: oddly, he spoke of how "we still have the team to reach the final".

- a. Eriksson and his team admitted losing on penalties again...
- b. Eriksson and his team admitted defeat on penalties again...
- c. Eriksson and his team finally had to admit defeat on penalties again...

was particularly tragic- "We practised penalties so much, I really don't know what more we could do about it", the Swede said.

Stimulus Set C: believe—luck

Hosts Germany turned out to be one of the positive surprises of the tournament. When their team was grouped with Costa Rica, Poland and Ecuador in last december's draw,

- a. the German fans did not believe their fortune in the draw...
- b. the German fans did not believe their luck in the draw...
- c. the German fans could hardly believe their luck in the draw...

would take them anywhere past the first knock-out round, and surely nobody expected Jürgen Klinsmann's team to beat an opponent like Argentina. Five matches into the cup it was 5-3 to Germany on penalties and it looked like nothing could keep them from storming into the final.

Notes

- * I am grateful to Anatol and Benjamin Stefanowitsch for programming the experimental software and supplying it to me. Many thanks also to Ewa Dabrowska, Stefanie Wulff and Kathryn Allan for running parts of the experiment. Earlier versions of this paper were presented at DGKL/GCLA 2 in Munich and ICLC 10 in Krakow. I would like to thank the audiences for discussion and useful comments. Finally, I am grateful for the very helpful suggestions of two reviewers for *Cognitive Linguistics* which have significantly improved the quality of the paper. All remaining errors are mine. Contact Address: Universität Bremen, Fachbereich 10, Postfach 33 04 40, 28334 Bremen, Germany. Author e-mail: zeschel@uni-bremen.de.
1. In the study by Garnsey and colleagues, SC-biased items were defined as verbs that occurred with sentential complements at least twice as often as with NP direct objects (as determined by a sentence completion task).
 2. Suitable items had to be both relevantly frequent and permit a substitution of the noun with a non-collocating close semantic variant for the collocationally neutral condition.
 3. Samples were evenly distributed across the corpus; samples for verbs reflected the proportions of the four different inflected forms in the overall concordance. The detailed figures are as follows (total corpus frequency in brackets): admit (11,283)–372 coded sample tokens; defeat (with noun tag: 3476)–346 tokens; believe (34,559)–380 tokens; luck (3180)–343 tokens; prove (14,593)–374 tokens; worth (3194)–343 tokens; transitive construction (3,747,626)–384 tokens (see note 5).
 4. For the noun *worth* in *prove one's worth*, it was not possible to adopt this approach since the BNC tagging was unusually inaccurate here: as it turned out, the vast majority of occurrences of *worth* with a nominal tag were not in fact nouns but wrongly classified adjectives (uses of the type *X is worth Y*). In order to address this problem, the following procedure was applied: first, the proportion of nominal uses of *worth* was estimated by manually analysing a sample of 373 tokens out of the overall 12,381 occurrences of the word (17.96 percent), thus giving an estimated 2224 nominal instances of *worth* in the entire corpus. On the basis of this figure, it was then possible to calculate the number of nominal observations that had to be analysed in order to estimate the proportion of transitive object uses among these 2224 tokens (328 examples). Finally, I began to analyse the complete concordance for *worth* (all tags) until I had identified 328 nominal tokens and then assessed how many of these featured *worth* as the head of the direct object constituent in a transitive construction (106), a figure that was then extrapolated to the overall population (thereby giving 32.32 percent or 719 tokens).
 5. The number of transitive constructions in the BNC was estimated by drawing a sample of 384 verb tags, counting the number of transitive constructions in the sample (141) and extrapolating its proportion to the overall number of verb tags in the corpus (giving an estimated 3,747,626 out of 10,206,300 verb

tags in total). This is essentially the approach advocated by Stefanowitsch and Gries (2003) who bootstrap argument structure construction frequencies from verb frequencies, an intuitively appealing yet admittedly less than perfect operationalisation since there is not a 1:1 correspondence between verb tags and argument structure constructions (for instance, an expression like *I really shouldn't have eaten that* manifests a single transitive construction, yet contains three items that would receive a verb tag in the BNC). While certain refinements of this measure would have been easy to implement (such as an exclusion of all modal verbs), this would not have been possible for other problematic cases such as auxiliary uses of *do* and *have* or light verb uses of e.g. *go* in serial verb constructions. Since it would have been unprincipled to exclude only some of the potentially problematic cases, I simply took the total number of verb tags in the corpus. Coding-wise, constructions were counted as transitive iff the main verb occurred with two arguments and the second argument was a direct object that could be passivized. Mood was ignored, meaning that examples like *He was hit by a truck* were coded as transitive. Finally, examples were counted as transitive as soon as *the relevant clause was transitive*, regardless whether the sampled verb tag itself did in fact belong to the transitive main verb or to an auxiliary.

6. For instance, of the 58 hits for *admit/admits/admitted/admitting...(w5)...defeat* in which *defeat* is actually the direct object of the verb, 20 have the infinitive marker *to* (tagged "T01") in position L-1. However, a look at the *lexical* co-occurrence statistics shows that the marker *to* actually appears 25 times in this position, with five occurrences (20 percent) erroneously tagged as a preposition ("PRP") instead.
7. Subjects had to be drawn from this wide geographic range because it had proved difficult to find enough native speaker participants among students in Bremen alone at the time of investigation. However, since regional linguistic differences between subjects can be assumed to be irrelevant for the task at hand, this was not deemed problematic.

References

- Bybee, Joan
2006 From usage to grammar: the mind's response to repetition. *Language*, 82, 4, 711-733.
- Connine, Cynthia, Fernanda Ferreira, Charlie Jones, Charles Clifton and Lyn Frazier
1984 Verb frame preferences: Descriptive norms. *Journal of Psycholinguistic Research* 13, 4, 307-319.
- Cuetos, Fernando, Don C. Mitchell and Martin Corley
1996 Parsing in different languages. In: Manolo Carreiras, Jose Garcia-Albea, and N. Sabastian-Galles (eds.), *Language processing in Spanish*. Hillsdale, NJ: Erlbaum, 145-187.
- Elman, Jeffrey L., Mary Hare, and Ken McRae.
2004 Cues, constraints, and competition in sentence processing. In Tomasello, Michael, and Dan Slobin (eds.), *Beyond Nature-Nurture: Essays in Honor of Elizabeth Bates*. Mahwah, NJ: Lawrence Erlbaum Associates, 111-138
- Ferreira, Fernando, and Charles J. Clifton
1986 The independence of syntactic processing. *Journal of Memory and Language* 25, 348-368.

- Ford, Marilyn, Joan Bresnan, and Ronald M. Kaplan
 1982 A competence-based theory of syntactic closure. In Bresnan, Joan (ed.), *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press, 727-796.
- Frazier, Lyn
 1987 Sentence processing: A tutorial review. In Max Coltheart (ed.), *Attention and performance XII: The psychology of reading*. Hillsdale, NJ: Lawrence Erlbaum Associates, 559-586.
- Frazier, Lyn and Janet Dean Fodor.
 1978 The sausage machine: A new two-stage parsing model. *Cognition* 6, 291-325.
- Garnsey, Susan M., Neal J. Pearlmutter, Elizabeth M. Myers, and Melanie A. Lotocky
 1997 The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37, 1, 58-93.
- Goldberg, Adele E.
 2006 *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Hare, Mary L., Ken McRae and Jeffrey L. Elman
 2003 Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language* 48, 2, 281-303.
 2004 Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes*, 19, 2, 181-224.
- Jurafsky, Daniel
 1996 A Probabilistic Model of Lexical and Syntactic Access and Disambiguation. *Cognitive Science* 20, 137-194.
- Langacker, Ronald. W.
 2000 A dynamic usage-based model. In: Barlow, Michael and Suzanne E. Kemmer (eds.), *Usage-Based Models of Language*. Stanford: CSLI Publications, 24-63.
- Lorch, Robert F. and Jerome L. Myers
 1990 Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16, 149-157
- MacDonald, Maryellen C., Neil Pearlmutter and Mark Seidenberg
 1994 Lexical nature of syntactic ambiguity resolution. *Psychological Review* 101, 676-703.
- McDonald, Scott A. and Richard C. Shillcock
 2003 Eye movements reveal the on-line computation of lexical probabilities. *Psychological Science* 14, 648-652.
- Pickering, Martin J., Matthew J. Traxler, and Matthew W. Crocker
 2000. Ambiguity Resolution in Sentence Processing: Evidence against Frequency-Based Accounts. *Journal of Memory and Language* 43, 3, 447-475
- Rayner, Keith and Lyn Frazier
 1987 Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology* 39, 657-673.
- Sinclair, John

- 1987 Collocation. A progress report. In: Steele, Ross and Terry Threadgold (eds), *Language Topics. Essays in honour of Michael Halliday. Vol. 2*. Amsterdam: John Benjamins, 319-331.
- Sinclair, John
1996 The Search for Units of Meaning. *Textus* 9, 75-106.
- Stefanowitsch, Anatol and Stefan Th. Gries
2003 Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8, 2, 209-243.
- Stefanowitsch, Anatol and Stefan Th. Gries
2005 Covarying Collexemes. *Corpus Linguistics and Linguistic Theory* 1, 1, 1-46.
- Tanenhaus, Michael K. and John C. Trueswell
1995 Sentence comprehension. In: Miller, Joanne L. and Peter D. Eimas, (eds.), *Handbook of perception and cognition Vol. 11: Speech, language and communication*. San Diego, CA: Academic Press. 217-262.
- Tanenhaus, Michael K., Michael J. Spivey-Knowlton, and Joy E. Hanna
2000 Modeling thematic and discourse context effects on syntactic ambiguity resolution within a multiple constraints framework: Implications for the architecture of the language processing system. In: Pickering, Martin J., Charles Clifton and Matthew Crocker (eds.), *Architecture and Mechanisms of the Language Processing System*. Cambridge: Cambridge University Press, 90-118.
- Trueswell, John C.
1996 The Role of Lexical Frequency in Syntactic Ambiguity Resolution. *Journal of Memory and Language* 35, 4, 566-585.
- Trueswell, John C., Michael K. Tanenhaus, and Christopher Kello
1993 Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 528-553.
- Trueswell, John C., Michael K. Tanenhaus, and Susan M. Garnsey
1994 Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language* 33 285-318.