How universal is Episodic Grammar? : A test in German

Jonathan Mallinson: 10407456

April 14, 2014

1 Introduction

Endel Tulving states that the human mind has two type of memory systems, semantic and episodic (Tulving, 1972). Semantic memory refers to a person's word knowledge in the form of concepts which are related to each other. Whereas episodic memory is a person's memory of specific events, (episodes) created from personal experience. Unlike the majority of other parsers which only consider semantic memory or episodic memory, episodic grammar (EG) combines both semantic and episodic memory in a cognitively plausible manner.

Borensztajn and Zuidema (2011) were successful in using EG to rerank English sentence of the wall street journal, achieving a F1 score of 90.61. No work however has been done on the use of EG outside of English. Although English is a popular choice to parse, it is not representative of natural language as a whole, due to its rigid word order and lack of case markings (Givón, 2001). A better solution would be to test parsers on a wide variety of languages. The criteria that parser are able to parse all languages is particularly relevant to parser that aim to be cognitively plausible. The process in which all language is processed within the brain is universal therefore any model based on the cognitive process should also be universal, a cognitive parser should be language agnostic.

Within this paper I will explorer how suitable EG is in regards to parsing German, as well as discussing various potential pitfalls and difficulties EG may run into. German was chosen due to the availability of lexical and syntactic resources as well it exhibiting several features not found within English. Within section 2 I will provide an overview of how EG works before moving onto discussing features of German within section 3. Section 4 is dedicated to benchmarking EG on German. Finally section 5 comments on the interaction between these German linguistic features and EG.

2 Episodic grammar

Currently there exists two main approaches to syntactic parsing, a rule based approach and an exemplar based approach (Borensztajn and Zuidema, 2011). A number of parsers consist of a set of context free rules applied in a top down manner to create the syntactic structure. Within this paper I will consider this semantic memory, as the rules refer to an abstract relational knowledge. The exemplar based approach can be consider more episodic in nature as it consists of remembering linguistic constituents greater than rules. Data-oriented parsing (DOP) was one of the first paradigms to bridge these two approaches, where the productive units of DOP range from a single word to a complete syntactic trees (Bod, 1992). These approaches however make little to no effort to be cognitively plausible. EG on the other hand tries to maintain cognitive plausibility, this is achieved by holding four assumptions about memory (Borensztajn and Zuidema, 2011):

(1) All episodic memories leave a physical memory traces in the brain. (2) Traces that belong to the same episode must bind together to create a sequence of semantic elements. (3) Semantic memory can trigger an associated episodic memory by recalling the sequence of traces that belong to the same episode. (4) Its possible to disambiguate episodic traces that overlap with each other.

2.1 The model of Episodic Grammar

Episodic grammar consists of a network of syntactic processing units (treelets), where each treelet represents a context free rule (S $\rightarrow$ NP VP). These treelets are then connected together if they appear within the same sentences, thus representing the episodic nature of EG. An example of the utterance "boy eats mangos" can be seen in figure 1. Formally a derivation of a sentence consists of a sequence of visits to treelets: $< t_1, t_2, ...t_n >$. To remember the order of the sequence correctly each treelet has two IDs (pointers) the unique sentence ID $< S >$ and the position within the derivation $< K >$. A derivation is the ordered list treelets visited. The position of each treelet is determined by the derivation strategy.

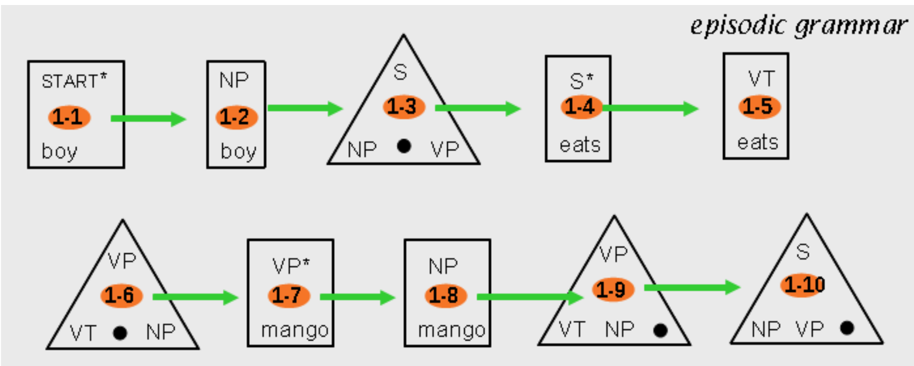


Figure 1: an EG trace of "boy eats mangos". Source: Gideon Borensztajn - lecture slides

The traced treelet can be used to parse seen sentences by following the pointers. EG can also be used to parse novel sentences. The first treelet selected is determined by the the derivation strategy (Borensztajn and Zuidema, 2011). This treelet then trigger memories of stored exemplars, where these exemplars will be any derivation which have the current treelet in common. The activation strength of the exemplars is determined by how similar the current derivation is to the exemplar. The exemplars compete against each with the the chosen transition strategy determining which exemplar to follow. With the exemplar chosen, the next treelet is retrieved from it and the process repeats.

The training of EG is summarized in the the following three steps (Borensztajn and Zuidema, 2011):

1. Creates a treelet for every unique context free rule seen in the treebank.
2. The order of each treelet needs to be determined $< t_0, t_1, ...t_n >$, this is done by the derivation strategy.
3. for every treelet leave a marker for the sentence ID and the position within the sentence $< s, k >$.

Once trained EG can be used to parse, as mentioned previously there are two transition strategies, a probabilistic EG and a shortest path EG. The shortest path approach is based on the idea of following the path of least resistance. The parser activation is determined by the length of the shortest derivation. In contrast the probabilistic approach basis the transition on how similar the two derivations are, as determined by their common histories. For the experiments within section 4 the probabilistic approach was taken.

In addition to the transition strategies there are several derivation strategies to choose from, including left corner parsing and top down approach. Within Borensztajn (2011) left corner parsing is favoured due to it achieving better parsing results than top down and its cognitive plausibility. Borensztajn (2011) argues that right corner parsing is more cognitively plausible as a study by Crocker (1999) demonstrated that incrementally within humans parsing is particularly salient.

3 Notable features of the German language

Although English and German share a common historical linguistic background, there are many features that are either not found in English or more prominently found within German. Some of these features are considered particularly difficult for parsers, including finite verb placement, flexible phrase ordering and discontinuous constituents (Kübler et al., 2006).

3.1 Finite verb placement

Within German the placement of finite verbs depends on the clause type (Kübler et al., 2006). In non-embedded assertion clauses the finite verb occurs at the second position of the clause (1). In yes/no questions the finite verb appears at the start of the clause (2). Within embedded clauses the finite verb appears at the end (3).

(1) peter wird das Buch gelesen haben. German (Kübler et al., 2006)
peter will the book read have
'Peter will have read the book'

(2) wird Peter das Buch gelesen haben. German (Kübler et al., 2006)
will Peter the book have read
'will Peter have read the book'

(3) dass Peter das Buch gelesen haben wird German (Kübler et al., 2006)
that Peter the book read have will
'that Peter will have read the book'

Variable verb placement is not a unique feature of German, it can also be found within many other languages, including Norwegian and Icelandic. (Thráinsson, 2010). Finite verb placement however is not found within English except for some minor exceptions coming from old and middle English (Kroch and Taylor, 1997).

3.2 Flexible phrase ordering

Although strict with the placement of finite verbs German is flexible in its phrase ordering. It is possible to move the complements and adjuncts of the main verb, As seen below (Kübler et al., 2006):

(4) Der Mann hat gestern dens Roman gelesen *German*[0]
the man has yesterday the novel read
'the man read the novel yesterday'

(5) Gestern hat der Mann den Roman gelesen *German*[0]
yesterday has the man read the novel

(6) Den Roman hat der Mann gestern gelesen *German*[0]
the novel has the man yesterday read

From this we can also see that verbs within German are left branching in contrast to right branching within English. Although the above demonstrates the freedom within phrase ordering, this freedom however is still limited compared to the free word order Slavic languages such as Slovene (Koktová, 1999).

3.3 Discontinuous Constituents

A third property of German syntax is the use of discontinuous constituents. Discontinuous constituents are constituents which cross each other within a syntactic tree, as seen within figure 2. Many linguistic theories of German posit that discontinuous constituents exist in both extraposed relative clause (7) and extraposed non-finite VP complement (8) (Kübler et al., 2006).

(7) Der Mann hat gestern den Roman gelesen, den ihm Peter empfahl.
the man has yesterday the novel read which him peter recommended

'Yesterday the man read the novel which peter recommended to him' - *German*[0]

[0](Kübler et al., 2006)

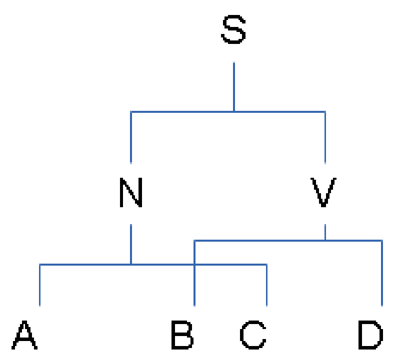


Figure 2: An example of a discontinuous tree

(8) Peter soll den Mann empfolhlen haben, den Roman zu lesen.
Peter is to the man recommended have the novel to read
'Peter is said to have recommended to the man to read the novel' - *German*[0]

German is not unique in its use of discontinuous constituents, many other language including English have linguistic theories that support the idea of discontinuity. In English they can be found within WH-fronting, Scrambling and Extraposition. WH-fronting is an example of long distance dependencies and involves the movement of WH word as seen in figure 3 (Lutz and Pafel, 1996). Scrambiling is a pragmatic word order where there is no set order for the words to appear in. This freedom can results in arguments of predicates crossing each other creating discontinuity. Extraposition allows for the movement of constituents to the right of their canonical position.

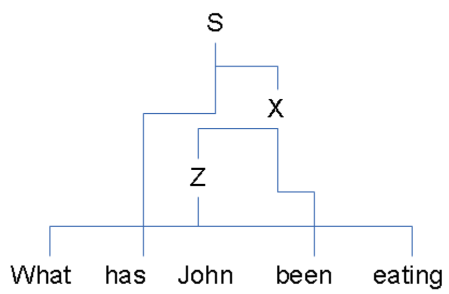


Figure 3: An example of WH-fronting

Syntactic constructs are latent structures resulting in evidence to show their existence being hard to come by. Therefore determining whether English speakers make use of discontinuous constituents is difficult. In German however there is more of a consensus over the matter and this consensus is reflected in the Negra treebank containing discontinuous constituents.

4 Experiment

To test the ability of EG to parse German I compared the results of EG to those of other German parsers trained on the same treebank. The chosen treebank was Tiger which contains over 1200 parsed utterances (Brants et al., 2002). The first 90% of the treebank will be dedicated to training with the remaining 10% being the testing set. Preprocessing involved removing all tree which contain a miss match between number of open and closed brackets. For the testing only those utterances with a maximum token length of 15 are considered.

Two parsers were evaluated, a standard non lexicalized right branching PCFG which implemented binarization and the same parser with vertical markovization. To these two parsers episodic grammar rereankers were used, assessing the best 10 and 20 derivations. The episodic grammar tried both left corner and top down derivation strategies. Under all tests lemmatization was disabled. The reranker remained largely the same as the one used within Borensztajn and Zuidema (2011) however modifications were made to the morphological processing to better fit German morphology. The changes included changing how unknown words were categorised, with the process now using German affixes and suffixes to categorize these words. As mentioned earlier EG was implemented as a reranker and not a parser, which is unfavourable as a reranker sets a performance cap and floor.

4.1 Results

The results of the experiment can be seen in figure 4. For each PCFG model a baseline reranker was include which randomly selected one of the 10 best derivations.

Parser	F1
PCFG	0.706
PCFG+EG-RB+10	0.743
PCFG+EG-TD+10	0.743
PCFG+EG-RB+20	0.736
PCFG+RBL	0.646
PCFG-Mark	0.691
PCFG-Mark+EG-RB+10	0.766
PCFG-Mark+RBL+10	0.652
PCFG-Mark+EG-RB+20	0.762

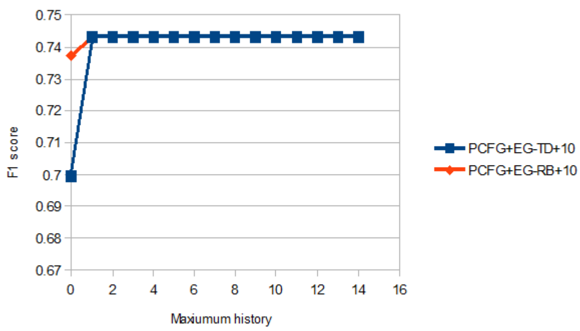
Figure 4: Key: PCFG, is the PCFG parser, PCFG-Mark indicates the parser with vertical markovization, EG is the episodic grammar, RB indicates the use of right branching derivation strategy, TD indicates top down derivation strategy, the number indicates the number of candidates to rerank, RBL indicates the random baseline

From the results we can clearly see that the EG model offered significant improvements over the PCFG models, in the case of the markovization model the improvement was over 6% and for the plain PCFG model it was 4%. All EG rerankers performed significantly better than the baselines proving the validity of the tests. Furthermore allowing the EG reranker the possibility of reranking 20 candidates did not lead to a significant difference on the results, indicating that EG was not reaching the cap on the maximum possible score.

The results show that EG is a valid approach in parsing German, as it offers improvements on not only the baseline PCFG model but the more advanced markovization model. However it does not prove that EG is universal neither does it prove that EG is able to parse German as well as English. For this to be tested two comparable treebanks would need to be tested. This test would require a parser and not a reranker to remove any bias introduced by the non EG parser. Even if the results were comparable then this process should be repeated for other languages including non Indo-European language. Although this process however may require the extension of the morphological implementation within EG.

What is interesting to note is that PCFG with markovization performed insignificantly worse than the plain PCFG, yet the EG reranker trained on this performed better than the PCFG without this information. This suggests that the PCFG is not able to make proper use of this additional information yet EG was able to.

We can also see that EG using right branching and top down approach score equally well. This is in direct contrast to the results reported by Borensztajn and Zuidema (2011). This however is not surprising as not all languages are right branching. As we saw in section 3.2 German verbs can be left branching, it is possible this affected the parser making right corner derivations less effective. German being well suited for left branching derivations is reflected in parsers trained for German taking a left corner derivational approach (im Walde, 2002). The advantage of EG is that it is possible to change the derivation strategy. If you take the Chomskyian universal grammar view of principle and parameters then you could state that left branching or right branching are simply principles which a language chooses and such the derivation should be language specific (Chomsky, 1995).



The above results show the effects of changing the maximum common history considered when determining the activation of competing derivations. In contrast to the results reported by Borensztajn and Zuidema (2011) on English, a history greater than one showed no improvements. This could be caused to the limited sentence length considered, further test should be conducted on less restricted sentence size to see if this is the case.

5 Feature analysis

5.1 Discontinuous Constituents

Discontinuous Constituents are problematic for EG, as these require that the grammar is at least as powerful as mildly context sensitive. This is not the case for EG which uses context free rules. There are two solutions to this problem either ignore discontinuous constituents or change the context free rules to a more powerful formalism. The first option was chosen for the above experiments where the treebank did not include discontinuous constituents. This approach is not ideal as both Swiss-German and Dutch are widely considered to be mildly context sensitive (Bresnan et al., 1987) (Shieber, 1987). As stated at the start, to make the claim of cognitively plausible the parser should support all languages.

linear context free rewrite systems (LCFRS) is a grammar formalism which is able to capture discontinuous constituents and offer many benefits. LCFRS are close to the semantic representation of an utterance, which is beneficial as this can be passed into a hierarchical prediction network in order to understand the semantics of the utterance (Borensztajn, 2011). LCFRS being mildly context sensitive is also advantageous as not only is natural language considered to be mildly context sensitive, but also as the expressive power of a language increases so does the complexity of parsing it (Kracht, 2003). Therefore it is ideal to use the grammar which is just able to understand natural language. LCFRS have also been successfully implemented within the DOP framework allowing EG to take advantage of the progress they have made(Van Cranenburgh et al., 2011).

LCFRS can be though of as a generalization of context-free grammar, where the rules are context-free (CFG) but instead of the rules consisting of strings, they consist of tuples(Van Cranenburgh et al., 2011). These tuples then consist of discontinuous symbols, instead of contiguous symbols as found within CFGs. The number of components in these tuples is called the fan out, it is equal to the number of gaps plus one. The fan out of the grammar is the largest fan out of all the rules.

Formally LCFRS is a tuple such that $G =< N, T, V, P, S >$. N is a finite set of non terminals. A function specifies the fan-out for every nonterminal symbol. T and V are finite sets of terminals and variables. S is the start symbol. P is the set of rewrite rules, which take the form of $s(a, b) \rightarrow NP_a VP_b$ or $s(ab) \rightarrow NP_a VP_b$, a coma indicates a discontinuity.

EG needs to changed into two major areas in order to support LCFRS. Firstly the treelet needs to be changed to representations rules of LCFRS. Secondly the derivation strategy needs to be changed. As the derivation strategy is modular the remaining model can be left unchanged. However it would be advantageous to change the binarization procedure.

Binarization techniques is a huge topic in itself, however in the case of LCFRS it is particularly important as it has a large effect on the processing time (van Cranenburgh, 2012). This is important as not only are LCFRS more complex than CFG but also EG is already resource intensive. Furthermore binarization determines the fan out, with larger fan outs being more complex. van Cranenburgh (2012) found that head driven binarization achieved the highest F1 score, therefore this technique should be chosen.

The most important aspect to change is the derivation strategy as left corner parsing is no longer applicable due to the non linear nature of LCFRS. One potential replacement candidate is Knuth's generalization of Dijkstra's shortest path algorithm as used by Kallmeyer and Maier (2010). Although left corner parsing was preferred due to its cognitively plausible approach, heuristic based approach may not be entirely implausible as many linguistic phenomenon take advantage of heuristics (Caramazza and Zurif, 1976) .

As there exists several treebanks containing discontinuous constituents these should be taken advantage of to test the performance of EG. These results can then compared to those of discontinuous DOP implementation. An interesting extension would be to test EG on English discontinues structure.

5.2 Flexible phrase ordering

The parsing of German could be hindered by the flexible phrase ordering. Flexible phrase ordering means that one semantic meaning could map to multiple utterances, as seen in section 3.2.

This is problematic for semantic memory based parsers as there are more syntactic rules to be learnt therefore more data needs to be provided to accurately determine the probability of these rules. It is also a significant problem exemplar based approaches as they must keep track of more examples. Below is an example of two utterance with the same semantic meaning (Note for simplicity I use English however the principle holds for German phrase ordering).

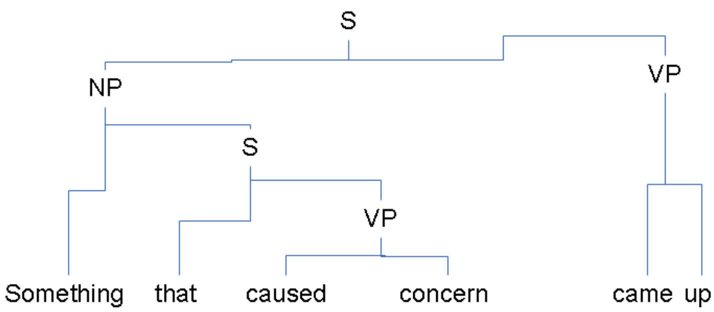


Figure 5: this is semantically equivalent to "something came up that caused concern "

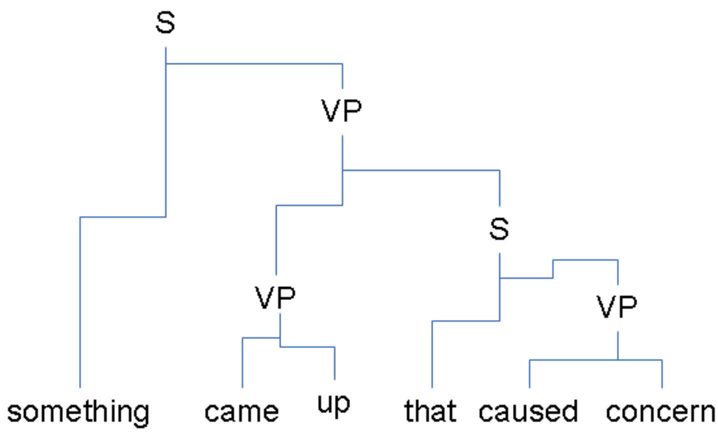


Figure 6: this is semantically equivalent to "something that caused concern came up"

From figure 5 and 6 we see that the two expressions that are semantically equivalent yet a left corner trace would be very different, where only two transitions would be the same between the two tree. We can also see that only two semantic rules are common between the two trees. This results in the EG and semantic parsers not being capable of fully generalising the two structures. This particularly bad for EG as it expends more resources than semantic parsers, yet would do equally badly learning from one tree and testing on the other.

The ideal solution to abstract away and realise that they both represent the same semantic structure. An LCFRS tree of these two utterances would be equivalent allowing for better generalization. This again shows the advantage of using LCFRS within an EG. It also further highlights how LCFRS allows for a better semantic structure, it is clear that if two tree share the same structure and lexical items then they must be semantically equivalent.

5.2.1 Experiment

To measure how problematic flexible phrase ordering is an experiment should be composed to check EG on its ability to parse it. Although it might be tempting to try and compare languages one with strict and one with flexible, this is difficult in practise because no two languages are different in just this regard, the results will be noisy. I suggest that two artificial grammars be created, one grammar should represent flexible word order and the other rigid. In order to test how well EG performs, EG should be compared against DOP , a plain PCFG and a state of the art parser such as the Stanford parser.

5.3 Finite verb placement

Finite verb movement can make it difficult for some parser may find it to generalise verbs behaviour Kübler et al. (2006). Again it is hard to see the effects of individual linguistic properties on parse results of natural language. Two artificial grammar should be constructed where one grammar contains finite verb placement and one does not, the test procedure should be the same as before. As EG is context sensitive and context sensitivity has been shown to improves parser performance(at a complexity cost), there is no reasons to suspect that EG would not do better than a conventional parsers (Bod, 1992).

6 Conclusion

From the results achieved in parsing German we see that episodic grammar offers significant improvements over both alternative parser. This suggest that yes EG is able to parse German. This is evidence towards the theory that EG is a universal grammar and therefore cognitively plausible. However further test on other languages need to be carried out to validate this claim.

As EG tries to be a truly cognitively plausible model it should focus on learning in a manner which resembles how children learn. Preliminary work has been done on EG and language acquisition, with the results being positive. However this should be continued and a natural extension would be an EG which can learn in an unsupervised manner.

References

Rens Bod. A computational model of language performance: Data oriented parsing. In *Proceedings of the 14th conference on Computational linguistics-Volume 3*, pages 855–859. Association for Computational Linguistics, 1992.

Gideon Borensztajn. *The neural basis of structure in language: bridging the gap between symbolic and connectionist models of language processing.* Institute for Logic, Language and Computation, 2011.

Gideon Borensztajn and Willem Zuidema. Episodic grammar: a computational model of the interaction between episodic and semantic memory in language processing. In *Proceedings of the 33th Annual Conference of the Cognitive Science Society*, volume 9, 2011.

Sabine Brants, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius, and George Smith. The tiger treebank. In *Proceedings of the workshop on treebanks and linguistic theories*, volume 168, 2002.

Joan Bresnan, Ronald M Kaplan, Stanley Peters, and Annie Zaenen. Cross-serial dependencies in dutch. In *The formal complexity of natural language*, pages 286–319. Springer, 1987.

Alfonso Caramazza and Edgar B Zurif. Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and language*, 3(4):572–582, 1976.

Noam Chomsky. *The minimalist program*, volume 28. Cambridge Univ Press, 1995.

Matthew W Crocker. Mechanisms for sentence processing. *Language processing*, pages 191–232, 1999.

Talmy Givón. *Syntax: an introduction*, volume 1. John Benjamins Publishing, 2001.

Sabine Schulte im Walde. A subcategorisation lexicon for german verbs induced from a lexicalised pcfg. In *LREC*, 2002.

Laura Kallmeyer and Wolfgang Maier. Data-driven parsing with probabilistic linear context-free rewriting systems. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 537–545. Association for Computational Linguistics, 2010.

Eva Koktová. *Word-order based grammar*, volume 121. Walter de Gruyter, 1999.

Marcus Kracht. *The mathematics of language*, volume 63. Walter de Gruyter, 2003.

Anthony Kroch and Ann Taylor. Verb movement in old and middle english: Dialect variation and language contact. *Parameters of morphosyntactic change*, pages 297–325, 1997.

Sandra Kübler, Erhard W Hinrichs, and Wolfgang Maier. Is it really that difficult to parse german? In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 111–119. Association for Computational Linguistics, 2006.

Uli Lutz and Jürgen Pafel. *On extraction and extraposition in German*, volume 11. John Benjamins Publishing, 1996.

Stuart M Shieber. Evidence against the context-freeness of natural language. In *The Formal complexity of natural language*, pages 320–334. Springer, 1987.

Höskuldur Thráinsson. Predictable and unpredictable sources of variable verb and adverb placement in scandinavian. *Lingua*, 120(5):1062–1088, 2010.

Endel Tulving. Episodic and semantic memory 1. *Organization of Memory. London: Academic*, 381:e402, 1972.

Andreas van Cranenburgh. Efficient parsing with linear context-free rewriting systems. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 460–470. Association for Computational Linguistics, 2012.

Andreas Van Cranenburgh, Remko Scha, and Federico Sangati. Discontinuous data-oriented parsing: A mildly context-sensitive all-fragments grammar. In *Proceedings of the Second Workshop on Statistical Parsing of Morphologically Rich Languages*, pages 34–44. Association for Computational Linguistics, 2011.