

Episodic grammar: a computational model of the interaction between episodic and semantic memory in language processing

Gideon Borensztajn (gborensztajn@uva.nl)

Willem Zuidema (zuidema@uva.nl)

Institute for Logic, Language and Computation, 904 Science Park

1098 XH, Amsterdam, The Netherlands

Abstract

We present a model of the interaction of semantic and episodic memory in language processing. Our work shows how language processing can be understood in terms of memory retrieval. We point out that the perceived dichotomy between rule-based versus exemplar-based language modelling can be interpreted in a neuro-biological perspective in terms of the interaction between a semantic memory system that encodes linguistic knowledge in the form of abstract rules, and an episodic memory that stores concrete linguistic events. We implement the idea of a semantic-episodic memory integration in a probabilistic grammar, and evaluate its performance as a syntactic parser on corpora of natural language. Our labeled precision and recall results are competitive with state-of-the-art syntactic parsers, with F-scores up to 90.68 on section 22 of the Penn WSJ corpus.

Introduction

Our ability to understand language depends, in part, on the declarative memory system, which is involved in representing factual knowledge and experiences with the external world. Declarative memory is, since Tulving [1972], further divided into two components: (i) *Semantic memory* is a person's general world knowledge, including language, in the form of concepts that are systematically related to each other; (ii) *Episodic memory* is a person's memory of personally experienced events or episodes, embedded in a temporal, spatial and emotional context. For example, the memory of the walk from your home to the bakery on a rainy Monday morning is part of the episodic memory system, while the concept of bread with all its associations is part of semantic memory.

Although there is general consensus that episodic memory and semantic memory are not separate modules, the exact nature of their relation is still an open question. A common view is that episodic memories are constructed as pointers that bind together items stored in semantic memory, both in temporal relations and in relations between roles or participants in an event [Shastri, 2002]. Such a conception of episodic memory fits with a popular theory in cognitive neuroscience, the so-called *reinstatement hypothesis of episodic retrieval*, which says that during episodic memory retrieval memory traces are triggered, thereby reactivating the cortical circuits that were involved in encoding the episodic memory [e.g., Woodruff et al., 2005]. Yet, to date there exists no theory of episodic memory that is sufficiently formalized to make quantitative and testable predictions about the relative contributions of episodic memory and semantic memory in any realistic cognitive task.

The distinction between episodic and semantic memory is also important for a cognitive approach to language research.

A substantial part of our memory is dedicated to representing linguistic knowledge in more or less abstracted form. The research we report here starts from the observation that the scientific debate on the relation between semantic and episodic memory can be meaningfully linked to an ongoing controversy about modeling language: one side in this controversy is focusing on empirical evidence for abstract, rule-based grammars [e.g. Marcus, 2001], while the other side emphasizes the item-based nature of grammar with a role (particularly in acquisition) for concrete sentence fragments larger than rules [e.g., Tomasello, 2000]. In our view, the rule-based grammars are best thought of as instances of *semantic memory*, since they encode abstract, relational linguistic knowledge. The item-based approach, on the other hand, suggests a role for *episodic memory* in sentence processing.

In computational linguistics Scha [1990] and Bod [1998] proposed Data-oriented Parsing (DOP) as a model that reconciles exemplar- and rule-based views. The productive units of DOP can range from single words to a complete phrase-structure trees. Such models however, as well as more recent Bayesian models [e.g., O'Donnell et al., 2009], operate on the *functional* level, but make no effort to include details from real cognitive or neural processes. The model we develop in this paper is also used for *parsing*: it assigns grammatical structure to new sentences, and disambiguates, using a probability model, between potentially very many grammatical analyses of each sentence. Assigning grammatical structure (rather than grammaticality judgments) as an intermediate step before semantic interpretation is, in this tradition, the essential function of grammar.

Unlike the vast majority of existing parsing models, however, we also aim to take some basic assumptions about episodic memory into account: (i) all experiences that occur in the lifespan of a human being, and that can be consciously remembered, leave physical *memory traces* in the brain. In the language domain, this means that every sentence ever processed will leave traces in memory; (ii) episodic memories are stored as temporal sequences of static semantic memories [e.g., Eichenbaum, 2004]; (iii) episodic memories are *content addressable*. This entails that their retrieval can be primed by cues from semantic memory.

In the next section we will give an outline of the model and discuss two different ways of representing earlier language experiences (corresponding to top-down and left-corner derivation). The ensuing section details how we collect the relevant statistics from a corpus of syntactically an-

notated sentences and define the probability model. In the remainder of the paper, we present our empirical results and discuss the relation between our approach and other work in statistical parsing and memory research.

Episodic grammar: model outline

In our approach we take the point of view that a grammar is instantiated in the brain as a network of interconnected autonomous syntactic processing units (henceforth *treelets*). For the sake of simplicity we assume in this paper that context-free rules (for instance $S \rightarrow NP VP$) from traditional grammars correspond one-to-one to such treelets. Treelets can serially combine with each other through dynamic binding (resulting in substitution), and they possess a register (internal memory) that keeps track of the correct order of application of the binding operations. This perspective on syntax generalizes to a connectionist account, which is worked out in [Borensztajn et al., 2009].

The network of interconnected treelets in our model constitutes a semantic memory of a grammar. We propose that the episodic memory of a sentence is distributed across the treelets, and consists of physical traces, contained inside the treelets, that keep a record of the treelet's participation in the derivation of the processed sentence. This is illustrated in Figure 1, which shows the episodic memory traces left behind in the network after processing the sentences *girl who dances likes tango* (orange traces) and *boy likes mango* (blue traces), corresponding to a top-down leftmost derivation strategy.

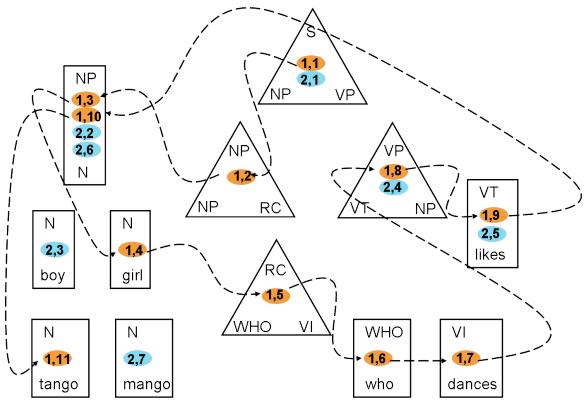


Figure 1: Episodic traces of two sentences (drawn as colored ovals) are stored in local memories of visited processing units (indicated by triangles and rectangles). Note, that by virtue of their ordinal number the traces implement *pointers* to successor treelets in a derivation (drawn for the first sentence alone)

Whereas in the context-free grammar framework a *derivation* of a sentence is a sequence of rule instantiations, in our framework, a derivation is a sequence of visits to treelets: $\langle t_0, t_1, \dots, t_n \rangle$. In order to remember the correct order of derivation (which can vary depending on the chosen derivation strategy) the episodic traces encode the sentence number (s) as well as the position (k) within a derivation. In Figure 1

the traces (for a top-down derivation) are identified by these two numbers, indicated as $\langle s, k \rangle$ inside the treelets. Note, that after hearing many sentences a single treelet will store traces for all sentences that have visited it, which are distinguished by their sentence number, and possibly multiple visits from the same sentence.

The episodic sentence memories distributed across the traces can also be recruited for the purpose of processing novel, unseen sentences. The idea is that when the derivation of a novel sentence arrives at a treelet, the traces encountered within the treelet trigger memories of stored exemplars. The strength of the activation of those memories depends on how much the stored derivations have in common with the current derivation (described more formally in the next section). Every next step in the derivation is determined by competition between traces of different exemplars, each having its own preference for a successor treelet, and its own activation strength. In this view sentence processing can be understood as subject to a *priming* effect: the traces prime or reactivate derivations of previously processed sentences (through content addressability), and restore the memory of previous parser decisions.

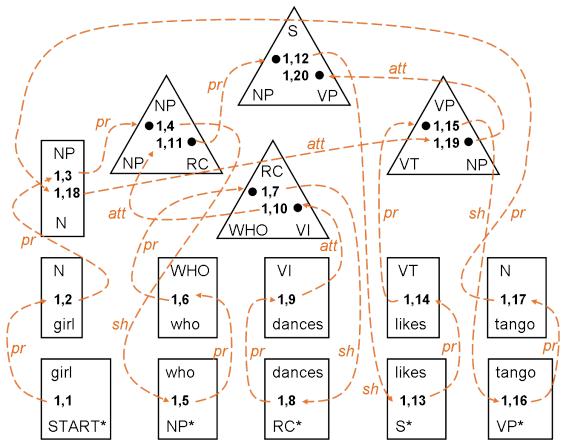


Figure 2: Episodic memory traces in the LCE grammar after deriving the sentence *girl who dances likes tango*.

One of the advantages of the episodic approach is that it allows comparison of different derivation strategies within a single framework, to find the effect of a different order of application of operations on treelets. An interesting parsing strategy from a cognitive point of view is *left corner parsing* [Rosenkrantz and Lewis II, 1970], since it proceeds incrementally from left to right and combines top-down and bottom-up parsing. Figure 2 shows an episodic left corner derivation for the sentence *girl who dances likes tango*. In left corner parsing treelets are introduced bottom-up by a *project* operation. As long as there are no *completed* (i.e., fully processed) treelets the next word in the sentence is introduced by a *shift* operation; otherwise the derivation can either *project* to a new treelet, or *attach* to a not yet completed treelet that has been previously introduced (In figure 2 these operations

are abbreviated with *sh*, *pr* and *att*). Note, that we have introduced special treelets that execute the shift to the next word (e.g., $RC^* \rightarrow \text{dances}$). These treelets employ *starred* non-terminals (e.g., RC^*): one or more stars indicate the register position in the treelet from where the shift operation originates (e.g., $RC \rightarrow \text{WHO}^* \text{ VI}$)¹.

One important difference with the top-down derivation strategy is that upon every attach operation treelets are reengaged in the derivation. It is therefore necessary to associate an episodic trace with a *specific register state* of the treelet, which keeps track of the operations (project, attach) performed on the treelet. This is indicated in Figure 2 by adding a dot before or after the trace².

Statistical Parsing with Episodic Grammar

To evaluate our model of episodic grammar we use a common task in the field of statistical natural language processing: disambiguation between competing syntactic analyses of a sentence (parses). Probabilistic grammars assign probabilities to different parses of a sentence and, in the most common setup, select the most probable one. One can estimate the parameters of a probabilistic grammar from a *treebank*, which is a corpus consisting of natural language sentences manually annotated with phrase structure trees.

After deciding on a derivation strategy (i.e., top-down or left-corner), the episodic grammar is trained by distributing a trace $e = \langle s, k \rangle$ in every visited treelet t_k of derivation $x = \langle t_0, \dots, t_k, \dots, t_n \rangle$ of sentence $\#s$ in the treebank. Specifically, given a treebank, the model:

1. creates an empty treelet for every unique context free production extracted from the treebank.
2. determines, for every treebank parse, the sequential order of treelets $\langle t_0, t_1, \dots, t_n \rangle$, according to the chosen derivation strategy.
3. leaves, for every step k in the derivation of sentence $\#s$, a trace in the visited treelet, encoded as $\langle s, k \rangle$.

Once all relevant statistics have thus been gathered, we can use the model to assign probabilities to candidate parses of a new sentence. Given an ongoing derivation d that has arrived at a certain treelet t_q , we define the probability of continuing the derivation to any other treelet $t_{q'}$ based on the activation values of the episodic traces stored in treelet t_q . The activation $A(e_{x_i})$ of the trace e_{x_i} (in t_q) of earlier derivation x is a function of the common history $\text{CH}(e_{x_i}, d)$ of derivation x (of which e_{x_i} is the i^{th} trace) with the ongoing derivation d . The CH is given by the number of derivation steps (i.e., treelets) that the stored derivation x and the pending derivation d have shared the same path before arriving at t_q . Episodic traces that share a long common history should contribute relatively

¹The addition of *shift treelets* allows for computing sentence probabilities, and thus turns the grammar into a language model. This approach is similar to the left corner parser of van Uyt sel et al. [2001].

²There can be as many register positions as there are children in the treelet.

much to the parser decision. A convenient choice for the activation of a trace is

$$A(e_{x_i}) = \alpha^{\text{CH}(e_{x_i}, d)} \quad (1)$$

where α is a parameter of the model. Depending on the chosen derivation strategy (e.g., top-down or left corner) the traces have different CH 's, hence receive different activations. All information to calculate these activations is stored inside treelet t_q ; computations are thus local, in line with our desire to define a cognitively and neurally plausible model.

The probability of moving to $t_{q'}$ in the next step of the derivation is simply the sum of activations of traces that point to $t_{q'}$, divided by the sum of all activations. Let $E_{t_q}^{t_{q'}}$ be the set of traces in treelet t_q that point to treelet $t_{q'}$, and E_{t_q} the full set of traces in treelet t_q . Then, the probability of moving the derivation to treelet $t_{q'}$ is

$$P(t_{q'}|t_q) = \frac{\sum_{e \in E_{t_q}^{t_{q'}}} A(e)}{\sum_{e' \in E_{t_q}} A(e')} \quad (2)$$

The probability of a complete derivation D is given by:

$$P(D = \langle t_0, t_1, \dots, t_n \rangle) = \prod_{i=1}^n P(t_i|t_{i-1}) \quad (3)$$

This probability can be computed dynamically, while simultaneously updating the common histories (and activations) of all traces at every step of the derivation. Let t_q and $t_{q'}$ be two successive treelets in the pending derivation d , and let $e' = \langle s, j \rangle$ be a trace stored in $t_{q'}$. Then its CH is updated according to

$$\text{CH}(e', d_{q'}) = \text{CH}(e, d_q) + 1 \quad (4)$$

if there exists a trace $e = \langle s, j-1 \rangle$ in t_q (i.e., a predecessor of e'). Otherwise, $\text{CH}(e', d_{q'}) = 0$.

In order to obtain a non-zero parse probability for test set derivations that contain words or productions not observed in the training set, we use standard smoothing techniques³.

The final step is to provide the episodic grammar model with candidate parses, that are assigned a probability according to equation (3). Since we don't have a specialized parser, but only a probability model, we use our model as a *reranker*. This entails that one takes a list of n best parses for every sentence produced by a third party parser (in this case Charniak's

³For unknown words we adopted the approach of [Sangati and Zuidema, 2011], replacing words that occurred less than 5 times in the train set by a class label. The treebank parses were binarized and horizontally Markovized, as proposed by [Klein and Manning, 2003]. Further, three levels of linear interpolation smoothing were used. The first level backs off to a non-episodic version of the used derivation strategy (e.g., PCFG probabilities), the second level backs off to less specific non-terminal labels, and the last level assigns uniform probabilities to all possible combinations of non-terminals that form unary and binary productions. These levels were parametrized by back-off parameters λ_1 , λ_2 and λ_3 . For technical details please refer to the webpage accompanying this paper: <http://staff.science.uva.nl/~gideon/cogsci2011>.

maximum entropy parser [Charniak, 1999]), and reranks the list by assigning a probability to each parse under the model of interest. We then used the PARSEVAL metric to evaluate labeled precision (LP), labeled recall (LR) and their harmonic mean (F-score) of the parses that receive the highest probability under the reranker [Manning and Schütze, 2000, p. 432].

Reranking does have some limitations as an assessment of the model’s performance, since the n best parses list produced by the third party parser has upper and lower bound precision and recall scores. For comparison we give the scores of a random reranker, that selects a parse from the list by chance. Confidence in the results of the reranker increases with the size of the n -best list (e.g., see Figure 4).

max his	top down reranker			left corner reranker		
	LR	LP	F	LR	LP	F
0	87.11	90.01	88.54	87.93	90.31	89.10
1	89.53	90.27	89.90	89.35	90.22	89.79
2	89.64	90.23	89.94	89.49	90.30	89.89
3	90.15	90.45	90.30	89.64	90.43	90.04
4	90.15	90.39	90.27	89.79	90.53	90.16
5	90.27	90.45	90.36	89.91	90.63	90.27
6	90.23	90.41	90.32	89.96	90.58	90.27
7	90.19	90.37	90.28	90.13	90.76	90.44
8	90.09	90.21	90.15	90.32	90.90	90.61
9	90.14	90.27	90.20	90.29	90.84	90.56
10	90.03	90.16	90.09	90.23	90.79	90.51
11	89.98	90.14	90.06	90.10	90.74	90.42
12	89.91	90.11	90.01	90.07	90.67	90.37
Ran		88.15	87.89	88.02		
Ch		90.23	90.15	90.19		

Table 1: Precision and recall scores of the episodic *top-down* reranker (columns 1-3) and *left corner* reranker (columns 4-6) as a function of the maximum history considered ($nBest=5$; $\alpha=4$; $\lambda_1=\lambda_2=\lambda_3=0.2$).

Experiments and results

We evaluated the performance of the episodic grammar on parsing the Wall Street Journal (WSJ). As is standard in statistical NLP, we used WSJ sections 2-21 for training. We used sentences of length upto 40 words from section 22 for testing (we reserve section 23 for evaluation of future versions of the model). The precision and recall results of the episodic top-down reranker, applied to the top 5 Charniak parses, are given in the first three columns of Table 1 as a function of the maximum common history that is taken into account by the episodic grammar (the column *max his*). CH’s larger than the maximum history are thresholded, resulting in maximum values for the trace activation, according to equation 1. The bottom 2 rows give the the scores for the random reranker and for our run of the Charniak parser. We have experimented with different parametrizations of α , λ_1 , λ_2 , λ_3 . Optimal results were obtained for $\alpha = 4$, and $\lambda_1, \lambda_2, \lambda_3$ in the range between 0.1-0.3, with only little variance.

In Table 1 and Figure 3 one can see a clear effect of conditioning history, peaking at history 5 for the top-down reranker, and at history 8 for the left corner reranker (best scores are indicated in boldface). For histories 3-7 the

episodic top-down reranker surpasses the Charniak F-scores by a slight margin, and overall does much better than the PCFG reranker (corresponding to *history 0*) and the random reranker.

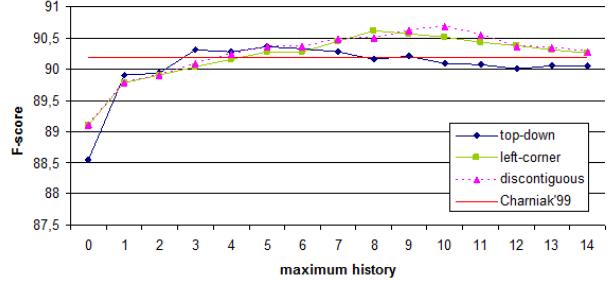


Figure 3: F-scores compared between the top-down and the left corner episodic reranker as a function of conditioning history. Also shown is a LCE reranker that uses discontiguities.

As can be seen from Table 1, the left corner episodic grammar (LCE) performs better across the board than the top down episodic grammar (TDE), and this is mainly due to improved labeled precision scores. It also does better than the probabilistic left corner model of [Manning and Carpenter, 1997], corresponding to *his= 0*. Note, that for the LCE reranker the peak is reached at history 8, and the F-scores stay high until history 14; this could be an indication that the order of conditioning in an LCE derivation better approximates human sentence processing than in a TDE derivation. It is remarkable that the LCE grammar – in its clean implementation, without any tricks popular in statistical parsing (like “head annotation”) – robustly improves on the Charniak parser, while the latter is the result of a very non-trivial engineering effort.

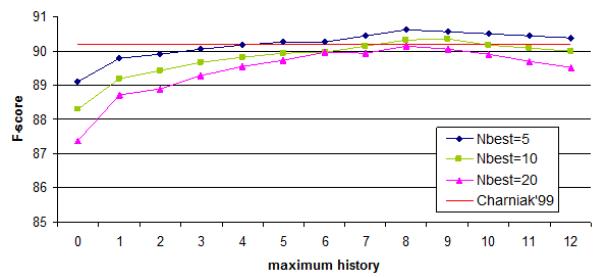


Figure 4: F-scores of the left corner episodic reranker applied to the top 5, top 10 and top 20 Charniak parses.

To assess the robustness of the reranking method we have also applied the LCE reranker to the top 10 and the top 20 lists of Charniak parses. In the latter case the random baseline is significantly lower than for the top 5 reranker (F-score = 86.2 resp. 88.0), but still the LCE reranker performs almost as good as Charniak (F-score=90.15 for history 8), while the top 10 reranker does even better (F-score=90.34 for history 9). In

Figure 4 it can further be seen that although the differences in performance between the top 5, top 10 and top 20 reranker are large for low histories, they converge for histories of 6-10, when the episodic approach starts to make a difference. Nevertheless, the fact that performance drops with larger n is a sign that our left-corner models are not yet capturing all relevant information usable for statistical parsing.

Discontiguous episodes An interesting way to extend the episodic grammar is by including *discontiguous* episodes. Our intuition is that language users often reuse memorized sentence fragments, even if they do not exactly match the sentence that is currently being processed. We implemented a variation of the LCE grammar that can exploit episodes with ‘gaps’. It stores discontinued episodes for later use, when it identifies a trace belonging to the same exemplar derivation as the interrupted episode. A fixed fraction f of the activation of the interrupted episode is then copied to the new trace. Best results were obtained when we let the activation of unused discontiguous episodes decay by some percentage d at every step of the derivation. With $d = 0.95$ and $f = 0.6$ the addition of discontiguous episodes gives a minor improvement over the base LCE model, as can be seen from Figure 3. The highest F-score is 90.68, which is reached for history 10. The effect of including discontiguous fragments seems to be that longer histories play an even more prominent role.

Shortest derivation reranker We have also implemented a shortest derivation LCE reranker that selects parses from the n -best list according to a preference for derivations that use fragments from the fewest episodes. With a best F-score of 90.44 (for history 9) it performs worse than the base LCE reranker, but still better than the Charniak parser.

Discussion

The role of sentence context in natural language processing (NLP) has in recent years seen renewed appreciation as is evident from the increasing popularity of statistical NL parsers that weaken in one way or another the context independence assumptions of probabilistic context free grammars (PCFGs). Context free grammars fail to take advantage of two relatively independent sources of contextual information for disambiguating between parses: *lexical context*, which captures the dependency on previous words in the sentence, and *structural context*, which captures the dependency on the relative position in a parse tree.

One remedy that is often used to cope with the lack of lexical context sensitivity of the PCFG is to *lexicalize* the grammar. Assuming that lexical dependencies are mostly carried between the head words of phrases and their dependents, one may enrich the constituent labels in the treebank trees with their head words, which are percolated up in the tree, and subsequently estimate the parameters of the PCFG from the lexicalized trees [e.g., Charniak, 1999]. The assumption of

independence of structural context, that is made by the standard PCFG model, is not very realistic either. For instance, in the WSJ treebank a noun phrase (NP) expands 9 times more often to a personal pronoun in subject position than in object position [Manning and Carpenter, 1997]. Johnson [1998] showed that the parsing accuracy of the treebank PCFG is greatly increased by incorporating structural information in the node labels, for instance by enriching the labels with the parent label.

In the episodic left corner grammar both lexical context and structural context are integrated in the conditioning history without any need for preprocessing of the labels. As such it is comparable to the tradition of *history based parsing*, which exploits the idea that the parser moves are conditioned on n previous parser decisions in the derivation history. A weakness of the latter approach is however that it leads to very large grammars and data sparsity, since all conditioning events are saved explicitly in equivalence classes. In the episodic grammar, parser decisions are conditioned on arbitrarily long histories, at no cost to the grammar size, because conditioning context is implicit in the representation, and is constructed explicitly only during on-line processing of a novel sentence. Since every exemplar is stored only once in the network, the space complexity of the episodic grammar is linear in the number of exemplars. Another difference with history-based parsers is that in the latter the association between the conditioning event and the sentence from which it originates is lost, whereas in the episodic grammar the identity of an exemplar that has contributed to a derivation step is preserved. Therefore, history-based parsers cannot make use of discontiguous episodes, but the episodic grammar can.

It is also interesting to compare the episodic grammar with Data Oriented Parsing (DOP) [e.g., Bod, 1998], which is a framework for exemplar-based statistical parsing. In DOP the primitive units of the grammar are not CF rules, but subtrees of arbitrary size, which are extracted from the parses of a treebank. In a certain sense DOP and episodic parsing are complementary: whereas in DOP the substitution of an arbitrary large subtree is conditioned on a single nonterminal, in the episodic parser the application of a context-free rule is conditioned on an arbitrary large episode. An advantage over DOP is that in the episodic framework a derivation can also be broken down into fragments according to other generative processes than top-down, for instance left-corner. This opens the possibility to utilise the episodic grammar as a language model in speech recognition. Moreover, in the episodic grammar it is not necessary to store every possible tree fragment explicitly. This is potentially an advantage over existing DOP implementations, which suffer from computational inefficiency due to very large grammars. The fact that stored episodes are automatically *reconstructed* from traces during the derivation of a novel sentence obviates a time-expensive search through an external memory (i.e., a treebank of fragments), and makes the episodic grammar *content-addressable*.

Hence, the episodic framework provides a promising new framework to unify traditional rule-based and exemplar-based approaches to (probabilistic) syntax within a single, neural perspective, and interprets them as semantic and episodic memory respectively. Table 2 shows how our results compare to state-of-the-art parsers. Note, that the latter are evaluated on section 23 of WSJ, while all our results are on section 22.

Various parser strategies (on WSJ sec 23)		
Parsing model	F (≤ 40)	F (all)
Charniak (1999) (max. entropy)	90.1	89.7
Petrov and Klein (2007)	90.6	90.1
Sangati & Zuidema (2011) (DOP)	-	87.8
Charniak and Johnson (2005) (reranker, $n = 50$)	-	91.0
This paper (on WSJ sec 22)		
TDE reranker ($n = 5$)	90.4	-
LCE reranker ($n = 5$)	90.6	90.1
LCE + discfg ($n = 5$)	90.7	-

Table 2: Comparison of the episodic reranker to state-of-the-art parsers, for sentences of length upto 40, or all sentences.

We see our work not only as an application in computational linguistics, but also as a contribution to memory research, which offers an explicit computational instantiation of the reinstatement hypothesis of episodic retrieval. We have proposed a representation of episodic memory, in which episodic memory is stored in distributed form *inside* local memories of associated semantic memory units. It can be conceived of as an imaginary life-long thread spun through semantic memory. These ideas are consistent with contemporary research in neuroscience, which emphasizes the construal of episodes in the hippocampus as contextually bound sequences of semantic memories [e.g., Eichenbaum, 2004]. The hippocampal model of Levy [1996] shows that during episodic sequence learning special ‘context neurons’ are formed that uniquely identify (part of) an episode. These may function as a neural correlate of the *counter* that we implemented in the traces. The episodic grammar model represents a first attempt to validate this theory of episodic memory within the language domain.

Future work

Parsing with episodic grammars looks like a very promising direction in both cognitive modelling and statistical parsing, of which we have only started to exploit the possibilities. We are currently developing a full episodic chart parser, that computes parse probabilities online via the principle of spreading activation from stored derivations to (the traces in) its states. This work is part of a larger project, which aims at modeling the episodic-semantic memory interaction within a connectionist framework, through integration with the HPN network [Borensztajn et al., 2009], resulting in a system that can learn syntactic treelets and their substitutability relations from unannotated text without supervision. We further intend to investigate syntactic priming effects, and recency effects in sentence processing.

Acknowledgements We thank Federico Sangati for technical assistance and Rens Bod, Stefan Frank and Remko Scha for useful comments. This research was funded by the Netherlands Organization for Scientific Research (NWO), through the Vici-project “Integrating Cognition” (nr. 277.70.006) to Rens Bod and the Veni-project “Discovering Grammar” (nr. 639.021.612) to Willem Zuidema.

References

- R. Bod. *Beyond Grammar: An experience-based theory of language*. CSLI Publications, Stanford, CA, 1998.
- G. Borensztajn, W. Zuidema, and R. Bod. The hierarchical prediction network: towards a neural theory of grammar acquisition. In *Proceedings CogSci’09*, 2009.
- E. Charniak. A maximum-entropy-inspired parser. Technical report, Providence, RI, USA, 1999.
- E. Charniak and M. H. Johnson. Coarse-to-fine n-best parsing and maxent discriminative reranking. In *Proceedings ACL’05*, pages 173–180, 2005.
- H. Eichenbaum. Hippocampus: Cognitive processes review and neural representations that underlie declarative memory. *Neuron*, 44:109–20, 2004.
- M. H. Johnson. Pcfg models of linguistic tree representations. *Computational Linguistics*, 24:613–632, 1998.
- D. Klein and C. D. Manning. Accurate unlexicalized parsing. In *Proceedings ACL’03*, 2003.
- W. Levy. A sequence predicting ca3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus*, 6:579–590, 1996.
- C. D. Manning and B. Carpenter. Probabilistic parsing using left corner language models. In *Proceedings ACL’97*, San Francisco, CA, 1997.
- C. D. Manning and H. Schütze. *Foundations of Statistical Language Processing*. MIT Press, Cambridge, MA, 2000.
- G. F. Marcus. *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. MIT Press, Cambridge, MA, 2001.
- T. J. O’Donnell, N. D. Goodman, and J. Tenenbaum. Fragment grammars: Exploring computation and reuse in language. Technical report, 2009. MIT-CSAIL-TR-2009-013.
- S. Petrov and D. Klein. Improved inference for unlexicalized parsing. In *Proceedings NAACL HLT’07*, pages 404–411, 2007.
- D. Rosenkrantz and P. Lewis II. Deterministic left corner parsing. In *11th Annual Symposium on Switching and Automata Theory*, pages 139–152, New York, 1970. IEEE Press.
- F. Sangati and W. Zuidema. A recurring fragment model for accurate parsing: Double-DOP. (under review), 2011.
- R. Scha. Taaltheorie en taaltechnologie: competence en performance. volume 11 of 7–22, pages 409–440.
- L. Shastri. Episodic memory and cortico-hippocampal interactions. *Trends in Cognitive Science*, 6(4), 2002.
- M. Tomasello. The item-based nature of children’s early syntactic development. *Trends in Cognitive Science*, 4(4):156–163, 2000.
- E. Tulving. Episodic and semantic memory. In E. Tulving, editor, *Organization of memory*. Academic Press, New York, 1972.
- D. van UytSEL, F. van Aelten, and D. van Compernolle. A structured language model based on context-sensitive probabilistic left-corner parsing. In *Proceedings of NAACL’01*, 2001.
- C. C. Woodruff, J. D. Johnson, and M. D. Rugg. Content-specificity of the neural correlates of recollection. *Neuropsychologia*, 43, 2005.