The Syntagmatic Paradigmatic Model: A distributed instance-based model of sentence processing

Simon Dennis
Centre for Human Factors and Applied Cognitive Psychology
The University of Queensland
Brisbane, Queensland 4072, AUSTRALIA
sdennis@humanfactors.edu.au

Michael Harrington
Linguistics Program, EMSAH
The University of Queensland
Brisbane, Queensland 4072, AUSTRALIA
m.harrington@mailbox.uq.edu.au

Abstract

The Syntagmatic Paradigmatic (SP) model is a distributed, instance-based account of sentence processing. Built on the Minerva II model of episodic memory (Hintzman 1988), it characterizes sentence processing as the retrieval of sets of associative constraints from long-term memory and the resolution of these constraints in working memory. In common with connectionist approaches, the SP model provides a data-driven account of language learning and does not make strong a priori assumptions concerning the nature of syntactic knowledge. However, the SP model overcomes the limitations of scalability and systematicity that have undermined the generality of neural network models of language phenomena. In this paper, the model is described and key properties are outlined.

1 Introduction

Connectionist models of language behaviour are highly attractive in that they provide a parsimonious account of both processing and learning. They are also able to derive grammatical class information directly from an input stream, and are naturally sensitive to statistics in the environment (Elman, 1993, Tabor & Tanenhaus, 1999). However, connectionist accounts also have two notable limitations that have restricted their wider acceptance. Current models, like Elman's simple recurrent network (SRN), are difficult to scale up to substantial portions of a language, both in terms of the size of the vocabularies they can accommodate and the number of grammatical structures they can capture. A second limitation has been the inability of connectionist models to account for the systematic nature of linguistic knowledge (Fodor & Pylyshyn, 1988; Marcus, 1998; but see Elman 1998, for counter arguments). These remain serious problems.

In this paper, we present a model of sentence processing and learning that maintains the advantages of the connectionist approach while addressing these two shortcomings. The Syntagmatic-Paradigmatic (SP) Model seeks to establish an instance-based account of sentence processing, using distributed representations, which characterizes sentence processing as retrieval from memory.

1. *An instance-based account of sentence processing*. Instance-based models have been successfully applied to a number of cognitive domains, including category learning (Smith & Minda, 1998), decision-making (Dougherty, 1999) and, more pertinently, to the language domains of phonology (Nakisa & Plunkett, 1998), lexical knowledge (Goldinger, 1998) and second language syntax (Williams, 1999). More recently instance-based models of sentence processing have also appeared (Daelemans, 1999). Unlike connectionist accounts, in which learning involves the gradual adjustment of a single set of weights, learning in instance-based models simply involves the inclusion of additional instances. The elimination of this critical bottleneck in connectionist accounts may allow instance-based models to scale to larger vocabularies and provide better coverage of the constructions available in a natural language.

2. *Using distributed representations*. However, current instance-based theories of sentence processing are symbolic. They are not capable of extracting grammatical class information from the input stream alone; being dependent instead on the availability of tagged (Argamon-Engelson, Daga & Krymolowski, 1999) or otherwise prepared corpora (Scha, Bid & Sima'an, 1999). By combining distributed and instance-based representations, the SP model combines the strengths of both approaches.

3. *That characterizes sentence processing as retrieval from memory*. The SP model focuses on understanding sentence processing and learning in terms of the memory processes involved. Sentence traces are stored and processing involves the retrieval of relevant traces, followed by the resolution of the word-by-word constraints embodied in these traces. A shift toward memory-based models in the sentence and text processing literature has been driven by the recognition of the importance of understanding the interface between memory and linguistic representation in language processing (McKoon & Ratcliff, 1998). Likewise, a serious attempt to understand the role of memory in processing may inform our understanding of representation and control issues in models of memory (Dennis, 2001a, 2001b).

In the following sections, we first describe the SP model and then outline some of its key properties.

2 Conceptual Framework of the SP Model

2.1 The Syntagmatic Paradigmatic Distinction

As the name indicates, the SP model makes a primary distinction between the *syntagmatic* and *paradigmatic* associations between words. Syntagmatic associations exist between words that often occur together, as in "green" and "grass¨. In contrast, paradigmatic associations exist between words that may not appear together but can appear in the same sentence context, as in "green" and "red". Studies in the early 1960s observed a shift in underlying lexical organisation from the syntagmatic to paradigmatic in the production of free associates. This shift was evident both as a function of development (Ervin 1961), and of training (McNeill 1966), but a satisfactory explanation of the shift was never forthcoming. This line of research was largely abandoned with the rise of generative grammar and the general disenchantment with associative models of language learning. However, many of the objections that were raised at this time, especially those related to the inability of stimulus-response links to handle long-term dependencies and a general lack of sensitivity to hierarchical structure, apply only to very simple associative systems (Elman, 1998). The SP model assumes that the same learning that underpins the syntagmatic paradigmatic shift also underpins language acquisition in general.

2.2 The Memory Traces

At the heart of the model are three kinds of long-term memory traces: lexical, syntactic and relational, each of which is defined in terms of syntagmatic and paradigmatic associations. A lexical trace is defined as the paradigmatic associates of a word across the corpus. A syntactic trace is the set of syntagmatic associations within a sentence. A relational trace is the set of paradigmatic associations within a sentence.

As an example, suppose the model had been exposed to the sentences set out in the Corpus in Table 1, "Steve kicks the can", “Dave kicks the ball”, etc. The lexical trace for Steve, in the context of this corpus, would be Steve ⇒ Bert, Steve ⇒ George, Steve ⇒ Dave, where ⇒ denotes an associative mapping. The syntactic trace for the “Alison is loved by Steve” would be {Alison ⇒ is, Alison ⇒ loved, is ⇒ loved, Alison ⇒ by, is ⇒ by, loved ⇒ by, Alison ⇒ Steve, is ⇒ Steve, loved ⇒ Steve, by ⇒ Steve}, and the relational trace for the target sentence would be {Alison ⇒ Ellen, Alison ⇒ Jody, Steve ⇒ Bert, Steve ⇒ George}. Note that although the lexical and relational traces both contain paradigmatic associations, the lexical trace is accumulated over the entire corpus for an individual word (e.g. in this tiny corpus *Steve* is bound to the distributed pattern containing *Bert*, *George*, and *Dave* while *Dave* is bound only to *Steve*), while the relational trace is a binding of the paradigmatic associations of each of the words in a given sentence. Note also that if one labels the set containing *Alison*, *Ellen*, and *Jody* as a “lover” role and the set containing *Steve*, *Bert*, and *George* as the “lovee” role, the relational trace is an extraction of the predicate information contained in the sentence. That is, the relational trace captures a form of deep structure (see section 3.5 Surface Structure Independence for further development of this point).

Table 1. The lexical, syntactic and relational traces in the SP model. Note ⇒ denotes an associative mapping. The comma separates mappings within a trace.

Corpus
Steve kicks the can.
Dave kicks the ball.
Ellen is loved by Bert.
Jody is loved by George.
Alison is loved by Steve.

Lexical Memory
Steve ⇒ Bert, Steve ⇒ George, Steve ⇒ Dave
Dave ⇒ Steve
Alison ⇒ Ellen, Alison ⇒ Jody
Ball ⇒ can
etc.

Syntactic Memory
Steve ⇒ kicks, Steve ⇒ the, kicks ⇒ the, Steve ⇒ can, kicks ⇒ can, the ⇒ can
Dave ⇒ kicks, Dave ⇒ the, kicks ⇒ the, Dave ⇒ ball, kicks ⇒ ball, the ⇒ ball
Ellen ⇒ is, Ellen ⇒ loved, is ⇒ loved, Ellen ⇒ by, is ⇒ by, loved ⇒ by, Ellen ⇒ Bert, is ⇒ Bert, loved ⇒ Bert, by ⇒ Bert
Jody ⇒ is, Jody ⇒ loved, is ⇒ loved, Jody ⇒ by, is ⇒ by, loved ⇒ by, Jody ⇒ George, is ⇒ George, loved ⇒ George, by ⇒ George
Alison ⇒ is, Alison ⇒ loved, is ⇒ loved, Alison ⇒ by, is ⇒ by, loved ⇒ by, Alison ⇒ Steve, is ⇒ Steve, loved ⇒ Steve, by ⇒ Steve

Relational Memory
Steve ⇒ Dave, can ⇒ ball
Dave ⇒ Steve, ball ⇒ can
Ellen ⇒ Jody, Ellen ⇒ Alison, Bert ⇒ George, Bert ⇒ Steve
Jody ⇒ Ellen, Jody ⇒ Alison, George ⇒ Bert, George ⇒ Steve
Alison ⇒ Ellen, Alison ⇒ Jody, Steve ⇒ Bert, Steve ⇒ George

2.3 The SP Architecture

In the SP model, sentence processing is characterised as the retrieval of associative constraints from long-term memory followed by the resolution of these constraints in working memory. Figure 1 shows the architecture.

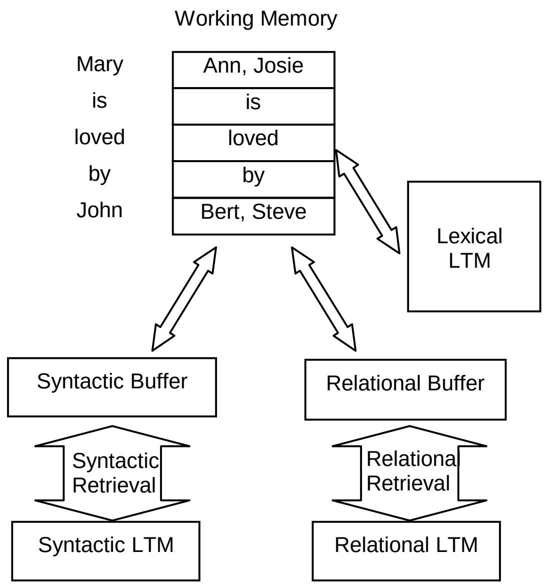

Figure 1: The SP Model architecture.

During interpretation, as each word of an input sentence becomes available, its physical form (either phonology or orthography) is used to retrieve its representation in lexical memory, that is, its distributed pattern of paradigmatic associates accumulated across the corpus. This pattern is inserted in the working memory slot adjacent to the input word. Note that at this stage the representation has not been influenced by the sentential context, so the pattern will contain both contextually relevant information as well as information that may eventually be excluded by the context. For instance, if the sentence were "The robber ran from the bank." the slot adjacent to "bank" might initially contain context relevant words such as "building", "institution" etc., but will also contain context irrelevant words like "shore", "edge", "hill", etc.

Next, the syntactic probe is formed by constructing the matrix of syntagmatic associations currently within working memory. This matrix is used to probe syntactic long-term memory, with the result returned to the syntactic buffer. The model uses the retrieved syntagmatic associations as a set of constraints on sentence interpretation and employs a gradient descent procedure to resolve these constraints. At this stage, sentential context will affect the patterns aligned with each of the input words. So, in the previous example, the context relevant words like "building" and "institution" will be reinforced by the constraints retrieved from memory, and context irrelevant words like "shore" and "edge" will be suppressed. That is, the model will have disambiguated the word sense. The model assumes that this process of word sense disambiguation is the same as that which is responsible for the assignment of thematic roles.

In a similar way, the relational trace is formed and used to probe relational long-term memory, with the result returned to relational working memory. The relational probe is constructed by forming the matrix of paradigmatic associations of each of the input words with what is currently in their adjacent slots. Again, the paradigmatic constraints in the relational buffer are resolved in working memory. This process allows known facts to influence sentence processing (and underpins how the model is able to answer questions, see section 3.6 Scaling below).[1]

[1] A more detailed technical account of the SP model is available at :

www .humanfactors.uq.edu.au/people/sdennis/

2.4 The retrieval model

Retrieval from long-term memory is achieved using a variant of the Minerva II memory model (Hintzman 1988). In Minerva II, the similarity of the probe vector is calculated for each of the traces in memory. The “echo” that is retrieved from memory is the sum of the trace vectors multiplied by their respective similarities. In the SP model the similarity measure used is the cosine of the angle between the probe and each trace raised to the third power. Figure 2 shows the Minerva II architecture.

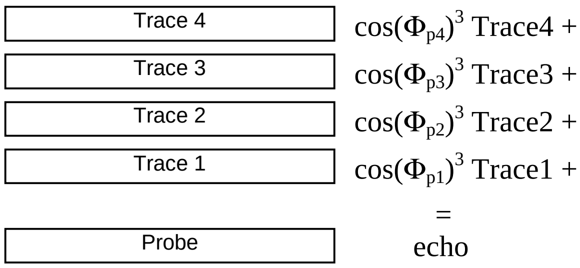

Figure 2: The Minerva II architecture.

The power varies, but is often set to three and is designed to make the most similar traces dominate the retrieved echo.

2.5 Constraint satisfaction

The final component left to explain is how the model adjusts the activations of words within the working memory buffer in order to satisfy the syntactic and relational constraints that have been retrieved from memory. In both cases a gradient descent procedure is employed.

If S_{ij} is the matrix of syntagmatic associations retrieved from memory then we minimize:

$$\sum_{ij} \left[S_{ij} - B_{ij} \right]^2$$

where

$$B_{ij} = \sum_{i=2}^{n} \left(\sum_{j=1}^{i-1} b_j \right) b_i^T$$

and b_i is the processing buffer vector in the i^{th} location. To update the buffer vectors the following equation is used:

$$\Delta b_k = \varepsilon \left[\sum_{l=1..k-1} b_l (S - B) + \sum_{l=k+1..n} b_l (S - B)^T \right]$$

where ε is a rate of change parameter and b_k are constrained to be positive. Resolution continues until:

$$\max(\Delta b_k) < C$$

where C is a parameter of the process.

Similarly, if R_{ij} is the matrix of paradigmatic associations retrieved from memory then we need to minimize:

$$\sum_{ij} \left[R_{ij} - B_{ij} \right]^2$$

where

$$B_{ij} = \sum_{i=1}^{n} b_i b_i^T$$

and b_i is the processing buffer vector in the i^{th} location.

To update buffer vectors, then, we use:

$$\Delta b_k = \varepsilon \, b_k (R - B)$$

Where ε is a rate of change parameter and b_k are constrained to be positive. Again, resolution continues until:

$$\max(\Delta b_k) < C$$

where C is a parameter of the process.

3 Properties of the SP Model

In this section, we will start by outlining how the SP model captures some of the important aspects of connectionist models of language processing, before demonstrating that it is capable of addressing some of their limitations.

3.1 Frequency

Like connectionist accounts, the SP model is inherently sensitive to frequency (or prior probability) information. Symbolic principle-based parser accounts (Frazier & Clifton 1996) predict that frequency will not influence early online parsing decisions. Although recent symbolic approaches have incorporated frequency information (Gibson 1998), they still assume a separate and substantial a priori syntactic knowledge base. By contrast, both frequency and structural information are derived directly from a single retrieval process in the SP model as described above.

3.2 Long Distance Dependencies and Structure Sensitivity

An early critique of simple associative models of language was that they were unable to account for long distance dependencies among words and phrases (Fodor, Bever & Garrett, 1974, e.g. The <u>governments</u> of the world <u>are</u> uniting against terrorism). Demonstrations by Elman (1993) have shown that associative models using a hidden layer can capture these phenomena. The SP model is also able to handle these structures. Instead of relying on a hidden layer to encode the relevant information, however, the SP model directly represents the dependency (using syntagmatic associations that span multiple words) and relies upon the sentential context to choose between stored traces containing the relevant associations.

To demonstrate the point, suppose that the syntactic memory of the SP model has been loaded with the following corpus:

1. The cat the cat sees sees.
2. The cat the cats see sees.
3. The cats the cat sees see.
4. The cats the cats see see.

Now suppose that the system is attempting to complete a sentence that begins with "the cat the cats". The syntactic probe created from this fragment would be {the $\Rightarrow$ cat, the $\Rightarrow$ the, the $\Rightarrow$ cats, cat $\Rightarrow$ the, cat $\Rightarrow$ cats, the $\Rightarrow$ cats}, which would match trace 2 most closely. As a consequence, the additional associations of: the $\Rightarrow$ sees, the $\Rightarrow$ see and see $\Rightarrow$ sees, would be added to the syntactic buffer. During resolution these constraints will ensure that the plurality information from the first noun is maintained over the embedded clause, and furthermore, that the clause structure is respected (i.e. "see" appears before "sees").

3.3 Generativity

In a model that operates by retrieving specific sentences instances, one may be concerned that only those sentence structures that appear in memory will be processed correctly, that is, that the model will be incapable of demonstrating the generativity that is characteristic of human language users. While it remains to be seen whether the SP model can capture the full scope of human performance, it is able to generalize beyond the specific instances in memory. Suppose we had the following corpus in syntactic memory:

1. the cat the man hears chases
2. the cat the men hear chases
3. the cats the man hears chase
4. the cats the men hear chase
5. the man the dog sees hears
6. the man the dogs see hears
7. the men the dog sees hear
8. the men the dogs see hear

and were trying to complete the fragment "the cats the men the dog". Now when we retrieve trace 4 and trace 7 will both be retrieved strongly. As a consequence both sets of syntagmatic constraints will be added to the syntactic buffer and the model will correctly complete with "sees hear chase", despite the fact that the model was never exposed to a sentence that contained two levels of center-embedding.

3.4 Systematicity

The strongest criticism of connectionist models of language processing and learning is that they are not able to capture the systematic nature of language (Fodor & Pylyshyn, 1988; Marcus, 1998; but see Elman 1998, for counter arguments). In particular, Fodor and Pylyshyn (1988) argued that you do not see people who can understand "John loves Mary", but who are unable to understand "Mary loves John." – that is, who can bind Mary to the object role but not the subject role. Connectionist models, however, do not automatically make this generalization and it has been argued that the amount of training data required in order to ensure such a generalization is unrealistic (Phillips 1994). By contrast, the SP model does exhibit such systematicity, because it is able to bind arbitrary role vectors to items when creating relational representations.

3.5 Surface Structure Independence

One of the ubiquitous characteristics of human languages is the way in which multiple surface forms can express the same underlying meaning. Traditional linguistic theories, especially in the generative tradition, have sought to capture this through a system of abstract formal rules. Within this tradition, it has been assumed that these rules must be innate and cannot be derived from inductive data-driven processes alone. Connectionist approaches, such as the SRN, have not directly addressed the issue and we believe will have significant difficulties in doing so, given the difficulty they have in capturing systematicity.

The SP model offers a mechanism by which relational structure can be captured in a surface form independent fashion. To illustrate the point assume the syntactic memory contained the following traces:

1. Bert loves Ellen.
2. George loves Jody.
3. Ellen is loved by Bert.
4. Jody is loved by George.

Now the relational representation of "John loves Mary" would be {Bert $\Rightarrow$ John, George $\Rightarrow$ John, Ellen $\Rightarrow$ Mary, Jody $\Rightarrow$ Mary}. However, the relational representation of "Mary is loved by John.", the passive form of the sentence above, is also {Bert $\Rightarrow$ John, George $\Rightarrow$ John, Ellen $\Rightarrow$ Mary, Jody $\Rightarrow$ Mary}. Despite the fact that the syntactic traces for the active and passive forms are very different, the relational traces are identical. That is, the model demonstrates surface form independence.

3.6 Scaling

Connectionist models are difficult to scale to substantive portions of a language, both in terms of the size of the vocabularies they can accommodate and the number of grammatical structures they capture (Elman 1993, Tabor & Tannenhaus 1999). Instance-based models, such as the SP model, do not encounter these scaling difficulties because learning occurs by appending instances to memory rather than applying a gradient descent optimisation procedure on a single set of weights, as is the case in connectionist models. In one demonstration, we presented the model with a set of short articles that were related to tennis and that appeared on the ABC news site in Australia. We then appended a set of short answer questions and answers to these articles, such as “Who defeated Pat Rafter in the Wimbeldon final? Goran Ivanisevic”. The model was able to answer such questions, on an article that was not part of the original corpus, showing that it can be applied to larger corpora with naturally occurring linguistic variation.

4 Conclusions and Further Work

The SP model represents a fundamental shift in the way we understand sentence processing and learning, and the role that memory plays in these processes. Current work is focused on verifying that the syntactic and relational priming that is predicted by the corresponding buffers in the model occurs in online reading tasks.

5 References

Argamon-Engelson, S., Dagan, I., & Krymolowski, Y. (1999). A memory-based approach to learning shallow natural language patterns. *Journal of Experimental and Theoretical Artificial Intelligence*, 11, 369-390.

Daelemans, W. (1999). Introduction to the special issue on memory-based language processing. *Journal of Experimental and Theoretical Artificial Intelligence, 11*, 369-390.

Dennis, S. (2001a). Generating distributed relational representations of sentences for information retrieval: Beyond the "Bag of Words". Twelfth Annual Winter Conference on Discourse, Text & Cognition, Jackson Hole, Wyoming.

Dennis, S. (2001b). The Syntagmatic Paradigmatic Model of Sentence Processing. Twenty-Sixth Annual Interdisciplinary Conference, Jackson Hole, Wyoming.

Dougherty, M. R. P. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106*(1), 180-209.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71-99.

Elman, J. L. (1998). Connectionism: Whence, Whither, and Why? Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society. Lawrence Erlbaum Associates.

Ervin, S. M. (1961). Changes with age in the verbal determinants of word association. *American Journal of Psychology*, 74, 361-372.

Fodor, J. A., Bever, T. G. & Garrett, M. F. (1974). *The psychology of language*. New York: McGraw Hill.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3-71.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1-76.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251-279.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95 (4)*, 528-551.

McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology, 49*, 25-42.

McNeill, D. (1966) A study of word association. *Journal of Verbal Learning and Verbal Behavior, 2*, 250-262.

Marcus, G. F. (1998). Can connectionism save constructivism? *Cognition, 66*, 153-182

Nakisa, R. C., & Plunkett, K. (1998). Evolution for rapidly learned representation for speech. *Language & Cognitive Processes, 13*(2/3), 105-127.

Phillips, S. (1994). Connectionism and systematicity. Proceedings of the Fifth Australian Conference on Neural Networks (ACNN'94), 53-56.

Scha, R., Bod, R., & Sima’an, K. (1999). A memory-based model of syntactic analysis: Data oriented parsing. *Journal of Experimental and Theoretical Artificial Intelligence*, 11, 409-440.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(6)*, 1411-1436.

Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science, 23*(4), 491-515.Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory and Cognition, 26*(4), 708-715.

Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition, 21*, 1-48.