Non-deterministic Prepositional Phrase Attachment

Jean-Pierre Corriveau
School of Computer Science, Carleton University
Colonel By Drive, Ottawa, CANADA
K1S 5B6
email: jeanpier@scs.carleton.ca

Abstract

Existing models of sentence comprehension typically adopt a deterministic approach that decides on *the correct parse* of a sentence. In essence, these models consist of algorithms that statically capture a *priori* rules for disambiguation and seldom take into account the context of interpretation. Also, their deterministic nature eliminates the possibility of recognizing a genuine ambiguity. We argue that the very essence of the quantitative model of memory we have developed, that is, its time-constrained nature, allows for a non-deterministic contextual approach to structural disambiguation. In this paper, we focus specifically on the problem of PP (Prepositional Phrase) Attachment. More precisely, we contend that a solution to this problem depends on the use of both a massively parallel time-constrained architecture and a quantitatively-defined context.

1. Introduction[1]

The problem of structural disambiguation stems from the fact that a sentence may have several parses. There are several facets to this problem. Hirst (1987, chapter 6), for example, partitions the possible structural difficulties of a sentence in two main categories: 1) analytic ambiguities and 2) attachment. Analytic ambiguities regroup several distinct kinds of decisions pertaining to the role of a syntactic unit in a sentence (*e.g.*, relative clause or complement, present participle or adjective, *etc.*) (*Ibid.*, p.149). The second category pertains to the attachment of a syntactic unit (*e.g.*, a **PP** (Prepositional Phrase), a relative clause, an adverb) to another syntactic unit of a sentence, as illustrated in the following examples (*Ibid.*, p.135):

PP attachment: to noun or verb?
1) Ross insisted on phoning the man with the limp.
2) Ross insisted on washing the dog with pet shampoo.

Relative clause attachment: to which noun?
3) the door near the stairs that had the ``Members Only" sign.

Structural ambiguity is not limited to these two categories and includes other hard problems such as gap finding (*Ibid.*, subsection 6.2.2). Furthermore, structural ambiguities can involve categorial ambiguities and can lead to a *garden-path* phenomenon, as in the following example where determining the end of the **NP** (Noun Phrase) is problematic (*Ibid.*, p.149) :

The prime number few.

It is often assumed that a sentence that presents a structural ambiguity generally has a preferred parse. However, Schubert (1986) remarks that a discussion of preferences often degrades to a battle of 'partisan informants', who often do not even agree on whether a sentence is 'confusing' or not. In other words, as with many other facets of linguistic comprehension, there is no well-accepted theory for structural disambiguation. For example, at the end of his analysis of the topic, Hirst (1987, section 6.3) remarks:

> There is at present no agreement on any general principles that can be used for disambiguation. It seems clear, however, that knowledge from several different sources is used.

Existing models of linguistic comprehension in general, and of structural disambiguation in particular, typically attempt to specify *the* rules for correctly interpreting a sentence, that is, for obtaining its *single determinate parse and meaning*. Such an approach seems problematic from a hermeneutic standpoint and we argue elsewhere (Corriveau, 1991a, ch.2) for a *reader-based* approach to interpretation. In this paper, we want to emphasize that a strictly deterministic approach to structural disambiguation presents two problems in that 1) it rules out the possibility of recognizing a genuine ambiguity and 2) it disadvantages a *strategic* view of comprehension (see van Dijk and Kintsch, 1983), that is, a model in which different heuristics (*e.g.*, for disambiguation) may be used at different points in time during interpretation.

[1] Support from the Natural Sciences and Engineering Coucil of Canada is gratefully acknowledged.

From a computational viewpoint, we remark that, generally, models for structural disambiguation are *qualitative* in that they specify complex rules, data structures and algorithms that involve conceptual entities. This leads to the sort of circular definition Minsky (1986, p.18) warns us against: cognition must be explained in terms of non-conceptual entities and processes. Selman's (1985) connectionist model for disambiguation constitutes an attempt at such a *quantitative* explanation. In the same vein, we have developed elsewhere (Corriveau, 1991a) a purely mechanical model of memory and have suggested that qualitative models for linguistic comprehension should be 'grounded' in such a purely quantitative architecture, that is, in an architecture that treats all conceptual knowledge as data. The fundamental and most pervasive hypothesis of our work is that, generally, linguistic comprehension is a *time-constrained* process--a *race*. Therefore, in our work, quantitative time, that is, processing time (henceforth simply time) as it pertains to memory management and memory processes plays a central role both as an organizational and an operational principle.

In this paper, we argue that the very essence of our quantitative model of memory, that is, its time-constrained nature, allows for a non-deterministic contextual approach to structural disambiguation. Due to space limitations, we will consider the specific problem of PP attachment. We contend that a solution to this problem, and to structural ambiguity in general, partly depends on the use of both a massively parallel time-constrained architecture and a quantitatively-defined context. We develop this thesis in the rest of this paper. We will start, in the next section, with a brief overview of existing computational models of structural disambiguation. We will then summarize, in section 3, the relevant characteristics of our model of time-constrained memory. Finally, in section 4, we will explain how this model of memory addresses the problem of PP attachment.

2. Existing Approaches To PP Attachment

It is not our intent to present here an exhaustive study of existing work in structural disambiguation nor in PP Attachment. We refer the reader to Hirst (1987) for a detailed discussion. Most of the studies on attachment decisions originate in the principles of *Right Association* and of *Minimal Attachment*, which are purely syntactic (McRoy, 1988, p.6):

> The principle of Right Association states that optimally, terminal symbols will be attached to the lowest non-terminal node that is on the right-most branch of the current structure; that is, they will be grouped with the terminal symbols immediately to their left.... Minimal Attachment... requires that optimally a terminal symbol is to be attached into a parse tree with the fewest possible number of new non-terminal nodes linking it with the nodes already in the tree.

The principle of Lexical Preference, which, in essence, states that lexical verbs and other lexical items may prefer one pattern of complementation to another, also plays an important role in recent discussions on attachment priorities. Against the syntactic trend, Wilks (1975, *et al.* 1985) argues for a more semantic version of lexical preferences, in which preferences correspond to selectional restrictions. Both Hirst (1987) and Schubert (1986) offer models that synthesize all these factors. For this reason, we will briefly overview only these two models.

Hirst suggests the use of a Semantic Enquiry Desk (SED), which is consulted by his parser for assistance with prepositional phrase attachment and gap finding in relative clauses. The SED requires:

1) An annotation on each verb sense as to which of its cases are 'expected' (COMPULSORY, PREFERRED, or UNPREFERRED).
2) A method for deciding on the relative plausibility of PP attachment.
3) A method for determining the presuppositions that would be engendered by a particular PP attachment, and for testing whether they are satisfied or not.
4) A method for resolving the issue when the strategies give contradictory recommendations.

Hirst acknowledges that deciding whether something is plausible is extremely difficult. Two strategies are taken to help somewhat. First, slot restriction predicates can be tested, failure indicating almost certain implausibility. Second, a knowledge base can be queried to test if it contains an instance of a given object, or an instance of something similar. If it does, then the object is judged to be plausible. A similar strategy is used for determining presuppositions. Finally, for contradictory recommendations, Hirst develops several ad hoc algorithms that ignore inferences, context, and pragmatics.

Schubert's approach to PP attachment is a lot more sketchy, involves numerically weighted preferences, and also allows for trade-offs among syntactic and semantic preferences. The model relies on the following six principles (1986, pp.601-602):

• **A graded distance effect**: Immediate constituents of a phrase prefer to be close to the *head lexeme* of the phrase. The effect is mediated by an 'expectation potential' that decreases with distance from the head lexeme and increases with constituent size; as a result, larger constituents admit larger displacements from the head lexeme.

• **A rule of habituation effect**: There is an inhibitory potential or 'cost' associated with each phrase structure rule (including lexical rules), leading to a preference for low-cost rules over high-cost rules.

• **Inhibition by errors**: 'Mild errors' such as concord errors contribute inhibitory potentials to the phrases in which they occur.

• **Salience in context:** The potential of a word sense or phrase is high to the extent that the denotation of that word sense or phrase is salient in the current context.

• **Familiarity of logical-form pattern**: The potential of a phrase is high to the extent that its logical translation instantiates a familiar pattern of function-argument combination.

• **Conformity with scripts/frames**: The potential of a phrase is high to the extent that it describes a familiar kind of object or situation (such as might be specified in a script or frame).

The first two principles are taken to capture syntactic preferences, and the others, semantic and pragmatic effects. In particular, the fourth principle "is intended to allow for semantic priming by spreading activation".

Closest to our work, McRoy (1988) proposes a psychologically plausible model of parsing that provides a good account of memory constraints, sentence complexity, structural preferences, and verb-frame preferences. The model employs a principled theory of grammar, concurrently processes syntax and semantics, and uses estimated timing information to resolve conflicting preferences. Also, McRoy agrees with us in viewing comprehension as a race process. However, in her work, the role of time is limited to attachment *hypothesizers* used to choose between conflicting preferences.

The fundamental feature of all of these conceptual approaches is that, contrary to neuronal models, they use symbolic data structures and algorithms to build a representation of the input. In contrast with the relative representational and processing arbitrariness of these models, the connectionist paradigm (Feldman, 1984) proceeds from the computational constraints imposed by neuronal modeling. In the last few years, considerable effort has been expended in order to develop natural language processing systems built on the massively parallel architectures and distributed processing that characterize connectionism. The resulting models are typically single-sentence parsers that produce a pattern of activation corresponding to a parse tree, as in Selman's (1985) work.

3. A Quantitative Architecture For Disambiguation

We root the quantitative cognitive architecture we propose in the basic metaphor of human memory. We develop at length elsewhere (Corriveau, 1991a) a strictly quantitative model of human memory 'on top of which' a conceptual analyzer for linguistic comprehension can be built. Of the multitude of possible factors that constrain comprehension, our architecture emphasizes the role and importance of *quantitative time*. The fundamental and most pervasive hypothesis of our work is that, in most situations, comprehension is a time-constrained process--a race. We summarize below the aspects of our model that pertain to structural disambiguation.

Time-constrained memory is partitioned into two temporal segments: a short term memory (STM) and a long-term memory (LTM). STM is assumed to have a limited capacity and its elements are taken to be, by definition, more readily accessible than those of LTM. Elements of the STM have a certain 'energy' (or activation) level that exponentially decays with time. Once the energy level of a member of STM falls below a certain threshold, it is either 'moved' to LTM or forgotten. Conversely, elements of LTM decay at a very slow rate over days or years--a process that is irrelevant from a computational perspective. The cognitive structures constructed by the process of comprehension in time-constrained memory are called *clusters* and 'move' between these partitions of memory. A cluster eventually decays out of STM into LTM, if it is not entirely forgotten. Also, the 'weakest' clusters in STM (with respect to their activation level) are the ones most likely to be 'moved' to LTM during memory management. Accessing a cluster is a time-constrained process, that is, a process that must complete its execution before a certain deadline. And a cluster is said to be *reachable* if and only if it can be accessed before this deadline.

LTM is taken to hold all user-specified qualitative data (*i.e.*, rules, schemas, *etc.*) in the form of a network of *knowledge units* (KUs), which can be thought of as static (as opposed to constructed) clusters. The perspective we take is that of the fundamental hypothesis of neuronal modeling (Feldman, 1984) and we adopt the consequent conceptualization of memory as a massively parallel network of simple computing units exchanging numeric signals. Memory processes (such as retrieval, forward and backward chaining, see Corriveau, 1991c) are constructed 'on top' of this network, and are taken to underlie the processes involved in cognitive tasks such as linguistic comprehension. The knowledge units in LTM at the beginning of an interpretation form the knowledge base to be used to construct new clusters during comprehension. The constructed clusters are organized in memory with respect to their reachability and constitute, at the end of an interpretation, the output of the model. Clusters are assembled from KUs, as well as from other clusters. In essence, a KU acts as an asynchronous feature[2] detector and cluster builder. A feature is a qualitative entity and its activation depends on the satisfaction a local constraint of one or more KUs. Each word of the text being processed activates some KUs, which in turn send numeric signals to other KUs. Because of the time-constrained nature of memory, a KU, once triggered by a signal, is given a small amount of time to accumulate enough energy (from the reception of other signals) to satisfy one of its constraints and become *activated* (or equivalently, *detected*). This is what we call a *detection race*. Once a KU satisfies one of its constraints, it can create and modify clusters in STM. In effect, the KU updates the contents of STM to reflect the activation of some feature(s). Thus, our model combines a connectionist-like architecture with the ability to *build* (rather than merely recognize) an interpretation.

In general, neuronal models are relaxation models (Selman, 1985): once inputs have been entered in the network, the computation terminates when the network reaches an equilibrium that is mathematically defined: a solution emerges from a constraint satisfaction process rather than from a search through a space of solutions. In contrast, conceptual approaches rely on data structures, rules, heuristics, and algorithms that must cover every possible input, and thus, in effect, define *a priori* the completeness of the understanding algorithm and the 'correctness' of the solution obtained. The idea of linguistic comprehension as a time-constrained process eliminates the need to choose between relaxation mechanisms and arbitrary rules to stop the understanding process. Instead, time becomes the essential stopping criterion for all memory processes And thus, the latter are non-deterministic in that they do not necessarily lead to a solution.

In our model, non-determinism is achieved by specifying KUs as mutually exclusive *candidates*. When candidates are simultaneously triggered by the activation of another KU (*e.g.*, from reading a polysemous word), we have a *candidacy race*: within a short amount of time either a winner must be chosen or some *resolution strategy* must be applied. Let us elaborate. Within the quantitative framework of time-constrained memory, we do not look for qualitative rules to choose between candidates, nor do we assume that a single candidate must be the winner. In fact, there could be several 'winning' candidates, a situation requiring the application of a resolution strategy. This strategy could decide to deactivate all candidates (leaving the input truly ambiguous), or it could invoke (*i.e.*, activate) a qualitative rule (*i.e.*, anoth KU) which, in turn, would choose the winne Several qualitative resolution strategies have bee suggested (*e.g.*, Granger and Holbrook, 1983). And has been stressed (*Ibid.*) that not only different reader may use different strategies, but also, that a read may use different strategies for different tasks an even for a same subtask at different points in tim (*e.g.*, because of a different context). From quantitative viewpoint, upon its activation, a K with mutually exclusive candidates sends a signal t all of the latter. We assume that KUs communica asynchronously and that the exchange of a sign between two KUs consumes time. Furthermore, w introduce the intuitive notion of *familiarity* to defin an *a priori* static order of retrieval for KUs in LTM More specifically, each KU has a user-specifie *retrievability coefficient*. If a KU sends a signal t two others, with the same communication delay fc both, then the most familiar of the two receivers, th is, the one with the lowest retrievability coefficien will receive the signal before the other one. Also, tw 'high priority' signals, namely 'forced activation' an 'forced inhibition', are assumed to have a priorit greatly superior to the other signals used by th system, that is, to 'instantaneously'[3] reach thei destination (regardless of the retrievability of thi destination). A candidacy race lasts a short amount c time and within this time-span, if only one winner i activated, it inhibits (using a forced inhibition signal all other candidates. As soon as two (or more winners are detected during a candidacy race, a KU responsible for a resolution strategy is activated. Thi KU, we repeat, can either inhibit all the specifie candidates or implement a qualitative resolutio strategy.

Finally, without going in details, we remar that, upon its activation, a KU may send a expectation signal to another KU. In this case, th receiver bypasses its normal constraint satisfactio process and becomes detected as soon as it receive another relevant signal.

4. Quantitative Aspects Of PP Attachment

As mentioned earlier, the problem of structura ambiguity includes several distinct facets of linguisti comprehension. In this paper, we restrict ourselves t the specific task of PP attachment. However, a explained in (Corriveau, 1991a), the mechanisms w present below appear to be relevant to other aspects c

2 A feature is a conceptual entity associated with one or more KUs.

3 'Instantaneous' communication, that is communication that requires zero time, is onl asymptotically possible and thus actually requires minimum time quantum in our model.

structural disambiguation. Furthermore, as a whole, structural disambiguation involves both quantitative and qualitative issues (Hirst, 1987). Within the framework of time-constrained memory, we will only address here the quantitative aspects of PP attachment. Also, in agreement with Wilks *et al.* (1985), we abandon the idea that there is necessarily a universal preferred parse for a structurally ambiguous sentence. Instead, we do believe it is desirable to have an approach that is non-deterministic, that is, that makes it possible for a sentence to be left unresolved because a genuine ambiguity is perceived. For example, consider the sentence:

Ross saw the man with the telescope.

in which the preposition 'with' could equally be attached to 'saw' and to 'man'.

Following the strategy we suggest for lexical disambiguation (Corriveau, 1993), the basic idea of our proposal is to treat the preposition of a PP as an ambiguous word. For example, in the above sentence, the word 'with' would be considered ambiguous (*e.g.*, withInstrument, withAttribute, withManner, *etc.*). The possible meanings of a preposition *prep* are specified as mutually exclusive candidates in a corresponding KU K_{prep}. Upon its activation, K_{prep} sends a signal to all its potential meanings. Each of these meanings corresponds to a separate attachment decision. Upon reception of this signal, each candidate attempts to become activated, independently of the others. In other words, there is no *a priori* discrimination against any of the possible attachments of the ambiguous preposition. In the above example, all interpretations of 'with' are, in theory, equally possible. However, several factors limit which attachment, if any, is chosen. First, because communication is taken to be asynchronous and considers the familiarity of the target receiver (see section 3), the signal sent by K_{prep} may not reach all candidate meanings at the same time. Therefore, not all candidacies necessarily start at the same time. Thus, it is possible that a candidacy race terminates before some of the candidacies have even started. In other words, familiarity, a notion we define quantitatively and that controls access to the candidates, implements a first form of preference. Second, expectations (which we take to subsume the phenomenon of semantic *priming*, as explained in Corriveau, 1991a) also quantitatively achieve a form of preference. More specifically, in the case where one of the possible interpretations of the ambiguous preposition is already being expected (from processing the previous words of the input text), the signal received from the preposition will immediately cause this candidate to become detected (possibly causing the inhibiting of other candidates). Such expectations may be syntactic in nature (*e.g.*, a particular verb admitting only a few specific prepositions could prime these) or may be contextual. Let us elaborate.

The notion of contextualized cognitive tasks is well accepted in psychology but has been typically downplayed in computational models. We take all cognitive tasks to occur with respect to a quantitative *context*, that is, with respect to a set of *reachable* features at a given point in time. Because reachability constantly varies as time elapses, so does context. Disambiguation is taken to consist in a set of concurrent time-constrained candidacy races. During these races, only reachable memory entities (*i.e.*, in context) are considered for constraint satisfaction. Thus, the context will directly affect whether a candidacy succeeds or fails. Also, because a candidacy race may span the processing of a few words (or even clauses) after the input of the ambiguous preposition, it is not limited to considering the context existing at its start, but in fact can take into account the context as it is modified over the time-span of this race. In other words, both the evidence existing at the start of a candidacy race and the information established from subsequent words can affect disambiguation, a highly desirable strategy. Furthermore, the time-constrained nature of the detection races implements *de facto* a decision point for the attachment. More specifically, unless a genuine ambiguity is detected (see section 3), the chosen attachment will be the one associated to the first candidate that 'wins'. And if all races expire without success, or if several races succeed, then the input will be left as ambiguous. Thus, the decision point for attachment is not specified in some qualitative algorithm but instead determined quantitatively by the time-span of the detection races: disambiguation must occur within the time interval defined by the time-spans of the different candidacy races.

To illustrate this discussion, let us consider the processing of the above example, which we consider truly ambiguous. The preposition 'with' triggers the candidates withInstrument, withAttribute, withManner, among others meanings. If the word 'saw' has led to the expectation of withInstrument, then immediately the latter is activated and may inhibit other contenders. Also, upon activation, withInstrument modifies the contents of STM so that the PP it governs becomes a subcluster of the cluster associated with the verb phrase. This, in effect, achieves the desired attachment. For simplicity, let us assume no such preference. In essence, each candidate implements a selectional restriction. For example, barring expectations, withInstrument can only become activated if it receives a signal from an instrument. Let us assume that the word 'telescope' leads to the satisfaction of the selectional restriction of both withInstrument and withAttribute, and thus to the activation of these KUs. Such multiple activations, in fact, lead to the application of a resolution strategy (see section 3) that will either deactivate the winners and leave the input ambiguous, or will attempt a qualitative heuristic. Context also influences attachment. For example, if the phrase 'the man with the telescope' refers to a previously processed man still in context, then the reference resolution process of our model (see Corriveau, 1991b) will solve this reference and short-circuit the candidacies of the meanings of with. In other words, contextual reference resolution can solve attachment.

To conclude, we observe that the model we propose for PP attachment shares some significant similarities with Schubert's models. These similarities are discussed in detail elsewhere (Corriveau, 1991a). Summarizing, we remark that time-constrained memory accommodates Schubert's principles, with the exception of the principle of the familiarity of the logical-form pattern, a clearly qualitative consideration. In essence, these principles can be rephrased in terms of time, races, expectations, and context. Hirst's work is definitely more qualitative (read algorithmic) and thus more distant from our research. However, we notice that his approach to relative plausibility and presuppositions relies on a reference resolution mechanism akin to the strategy we propose for the interpretation of definite references (Corriveau, 1991b). But, ultimately, we feel especially close to Small's (1980) *word expert parsing* philosophy in that we do not look for principles of disambiguation. Instead, each preposition triggers an expert (in the form of a candidate) for each of its possible interpretations and each noun and verb has the ability to specify, in its own expert(s), its particular preferences for attachment, by means of expectations.

5. Conclusion

The problem of structural disambiguation is crucial for linguistic comprehension. Within the non-deterministic framework of the quantitative model of memory we have proposed, disambiguation results in part from the existence in memory of a context, as well as from the fact that feature detection races are concurrent time-constrained processes. Furthermore, our model departs from traditional approaches it that it allows for the recognition of truly ambiguous input.

References

[1] Corriveau, Jean-Pierre (1991a) *Time-Constrained Memory for Reader-Based Text Comprehension*, doctoral thesis, department of Computer Science, technical report 246, University of Toronto, Canada.

[2] Corriveau, Jean-Pierre (1991b) "Interpretation of Definite Reference with a Time-Constrained Memory", *Proceedings of the Thirteenth Annual Conference of the Cognitive Society*, Chicago, IL.

[3] Corriveau, Jean-Pierre (1991c) "Time-Constrained Memory", *Proceedings of the first workshop on Parallelism and AI*, IJCAI-91, Sydney, Australia, pp.18-24.

[4] Corriveau, Jean-Pierre (1993) "Beyond Deterministic Lexical Disambiguation", submitted to: *PACLING 93*, Vancouver, Canada.

[5] van Dijk. T. and Kintsch W. (1983) *Strategies of Discourse Comprehension*, Academic Press, NY.

[6] Feldman, Jerome (1984) "Computational Constraints from Biology", *Proceedings of the Sixth Annual Conference of the Cognitive Science Society*, Boulder, Colorado, p.101.

[7] Granger, Richard and Holbrook, Jennifer (1983) "Perseverers, Recencies and Deferrers: New Experimental Evidence for Multiple Inference Strategies in Understanding", *Proceedings of the Fifth Annual Conference of the Cognitive Science Society*, Rochester, NY.

[8] Hirst, Graeme (1987) *Semantic Interpretation and the Resolution of Ambiguity*, Cambridge University Press, Cambridge, England.

[9] McRoy, Susan (1988) *The Influence of Time and Memory Constraints on the Resolution of Structural Ambiguity*, Technical Report CSRI-209, Dept. of Computer Science, University of Toronto, Canada.

[10] Minsky, Marvin (1986) *The Society of Mind*, Simon and Schuster, New York.

[11] Schubert, Lenhart (1986) "Are There Preference Trade-offs in Attachment Decisions?", *Proceedings of the National Conference on Artificial Intelligence*, AAAI, Philadelphia, PA, pp.601--605.

[12] Selman, Bart (1985) *Rule-based Processing in a Connectionist System for Natural Language Understanding*, Technical Report CSRI-168, Computer Science, University of Toronto, Toronto.

[13] Small, Steven (1980)*Word Expert Parsing: A Theory of Distributed Word-Based Natural Language Understanding*, Ph.D. Thesis, Technical Report 954, Dept. of Computer Science, University of Maryland.

[14] Wilks, Yorick A. (1975) "A Preferential Pattern-Seeking Semantics for Natural Language Inference", *Artificial Intelligence*, 6:53-74.

[15] Wilks, Yorick, Huang, Xuiming, and Fass, Dan (1985) "Syntax, Preference, and Right Attachment", *Proceedings of the Ninth Joint Conference on Artificial Intelligence*, Los Angeles, CA., pp.779--784.