Human Language Processing: Symbolic Models

S Vasishth, University of Potsdam, Potsdam, Germany
R L Lewis, University of Michigan, Ann Arbor, MI, USA

In this article we take 'symbolic models of sentence processing' to mean approaches in which processing is characterized in terms of relatively discrete, combinatorial representations. There are other properties of symbolic processing that are important for cognitive modeling more generally – for example, the ability of symbol processing to support Turing-general computation (Newell, 1990; Lewis, 2000a) – but for language processing, discrete combinatorial representations are often thought to be the hallmark of symbolic models. Defined in this manner, symbolic sentence processing models include approaches sometimes termed 'hybrid,' such as activation-based production system models, or localist connectionist models.

One important way to distinguish symbolic processing models is in terms of their ontological commitments, that is, the symbolic primitives out of which representations are composed. Arriving at an empirically and explanatorily adequate ontology of linguistic primitives is the traditional province of syntactic theory. Instead, we focus here on the variety of theoretical approaches to symbolic sentence processing that are largely orthogonal to the ontological distinctions made in syntactic theory – though often interacting with them in empirically and theoretically critical ways. We also focus here primarily on syntactic processing in comprehension, the domain of sentence processing that has received the most attention, though symbolic models exist that extend upward to discourse-level phenomena (Kintsch, 1988; Carpenter *et al.*, 1995).

The remainder of the article is organized around the following four broad approaches to symbolic sentence-processing theory, which jointly cover most existing models: (1) grammar-based approaches, which posit a tight link between competence grammars and processing phenomena; (2) symbolic complexity metrics and ambiguity resolution principles, which are abstract processing theories predicated on combinatorial syntactic representations; (3) approaches based on existing or independently motivated computational cognitive architectures; and (4) approaches based on probabilistic grammars, which appeal to frequency and experience, as encoded in a symbolic grammar formalism. Rather than surveying the entire literature, we illustrate each of the approaches by explaining the details of a few models, and provide pointers to the literature for other examples. We conclude the article with a brief section outlining points of commonality with connectionist models.

Grammar-Based Approaches

Chomsky (Chomsky, 1965) has long argued that any processing theory must include a theory of competence as a critical component, but given that a competence theory is a characterization of knowledge state that abstracts away from processing, it alone does not make processing predictions of the kind that bring it into contact with fine-grained psycholinguistic data: specific linking assumptions are required to bridge the gap. Grammar-based approaches presuppose a syntactic theory, and associate with it some fairly strong linking assumption so that the competence theory itself can be seen to drive the processing predictions.

The strongest form of this linking has been termed by Berwick and Weinberg as the type-transparency assumption. It states that "the logical organization of rules and structures incorporated in a grammar [is] mirrored rather exactly in the organization of the parsing mechanisms" (1984: 39). Many instances of this assumption can be found in the literature, and Berwick and Weinberg (1984: 39–82) provided an extensive discussion of the potential theoretical and empirical consequences of various positions.

In strong grammar-based approaches, the primitives of competence grammar are the principal explanatory force behind any observed processing difficulty. The metrics for determining processing complexity may differ as a function both of different underlying syntactic theories and of processor–grammar linking assumptions. Some theories rely on the history of the syntactic derivation and/or the number of basic operations involved (e.g., the derivational theory of complexity), while some rely on the number of rules used or violated (e.g., optimality theory).

The derivational theory of complexity (Miller and Chomsky, 1963; Fodor *et al.*, 1974) is the oldest such idea. Its central hypothesis is that the transformational model could be construed as a processing theory in addition to a competence theory (e.g., Chomsky, 1965) if "our performance on tasks requiring an appreciation of the structure of transformed sentences is some function of the nature, number and complexity of the grammatical transformations involved" (Miller and Chomsky, 1963: 481). Although early experiments showed little evidence to support such a connection between grammar and processing (e.g., Slobin, 1966), subsequent work has attempted to tackle the question in at least three distinct ways.

The Strong Type-Transparency Approach

One approach has been to redefine the architecture of grammar in the hopes of maintaining the strong type-transparency assumption. Examples are other versions of transformational grammar (Berwick and Weinberg, 1984; Phillips, 1990), lexical-functional grammar (Bresnan, 1978; Bresnan and Kaplan, 1982), and optimality theory (Fanselow *et al.*, 1999; Hoeks and Hendriks, 2004; Stevenson and Smolensky, forthcoming).

As an example of a sentence-processing model based on a new grammar architecture, we consider a specific approach based on optimality theory (OT) (Fanselow *et al.*, 1999). OT (Legendre *et al.*, 2000) is a grammar framework that assumes that there exists an evaluator, EVAL, and an ordered list of rules. The rules are violable, with the additional constraint – termed strict dominance – that no number of violations of a lower-ranked rule are as important as a violation of a higher-ranked one. Also provided is a function called GEN that takes as input a string (in sentence parsing, a string of words) and produces as output a set of so-called candidates. For example, given the set of words {*guess*, *we*, *saw*, *whom*} GEN would generate all the possible permutations. The candidates are simultaneously examined with respect to the ordered rules and a winning candidate declared based on the rule violations and subject to strict dominance.

This grammar architecture can be applied to sentence processing by relying on the following critical property of OT: the incremental, nonmonotonic application of the defeasible rule system allows for a ranking of contrasting sentence structures to be established, which predicts preferred parses. A simple example is the string {*guess*, *we*, *saw*, *whom*}. The two relevant rules whose relative ranking of candidates determine the outcome are antisymmetry (AS) and wh-criterion (whC); AS says that objects follow the verb, and subjects precede it, and whC that a constituent question must begin with a wh-phrase. Assuming that whC is ranked higher than AS, the winning candidate *guess whom we saw* violates only AS, whereas the permutations *guess we saw whom* and *guess we whom saw* violate the higher ranking whC constraint; and so on.

A potentially interesting property of OT is that it is considered a limiting case of a connectionist-inspired grammar formalism (Smolensky and Legendre, 2005); for example, the connectionist correlate to GEN is the start state of a network in which all possible constituents are represented as vectors, and the threshold logic units representing suboptimal candidates are deactivated, returning an optimal winner (Hale and Smolensky, 2001).

An Automata-Theoretic Approach

A second approach, championed by Joshi (Joshi, 1990; Rambow and Joshi, 1994), is to establish mathematical equivalences between grammar formalisms and computational automata, and then use processing complexity metrics naturally defined on these automata. Joshi presents complexity metrics over a (bottom-up) embedded pushdown automaton (BEPDA) to explain the difference in processing difficulty of, for example, Dutch vs. German center embedded structures (Bach *et al.*, 1986). The novel aspect of their automaton-based metric is that the BEPDA is related in a systematic way to the class of tree-adjoining grammars (TAGs), a framework for defining competence grammars (Abeillé and Rambow, 2000). The relationship is that for every TAG there is a BEPDA that accepts exactly the strings that the TAG generates, and for every BEPDA there is a TAG that generates exactly the set of strings that the BEPDA accepts. The uniqueness of this approach stems from the fact that formal models of competence and of performance are linked by the equivalence relation between the machinery; this mathematical relationship – rather than using a grammar formalism directly as a sentence-processing model – constitutes the link between grammatical representation and processing.

Grammatical Representations with an Information-Theoretical Constraint

The third type of exploration of the grammar–parser relationship is exemplified by the entropy reduction hypothesis (Hale, 2004). This model fixes a particular complexity metric (the reduction of entropy [Shannon, 1948] during incremental parsing) derived from a probabilistic grammar, and compares the effectiveness of modeling sentence-processing data with two distinct grammatical representational possibilities (Kayne's promotion analysis [Kayne, 1994] and Chomsky's adjunction analysis [Chomsky, 1977]). Although not an explicit goal of this approach, it provides a principled comparison between one set of representational assumptions vs. another in a precise manner. What makes this comparison possible is that the proposal is grammar independent and can be implemented for any grammar in the class of mildly context-sensitive grammars.

Some General Issues in Grammar-Based Models

In the grammar-based approaches, two general issues deserve comment. An assumption (either implicit or explicitly stated) in most grammar-based approaches is that the human parser is incrementally constructing syntactic structures and, furthermore, that building syntactic structure is a prerequisite for semantic interpretation (e.g., Frazier and Clifton, 1996). But an interesting development in linguistic and psycholinguistic theory is the possible elimination of the distinction between syntactic structure building and semantic interpretation. In particular, the concurrent, incremental construction of a combined semantic and syntactic structure has emerged as an alternative position (Steedman, 2000). This has important implications for modeling real-time sentence-processing data, since considerable evidence now points to a simultaneous use of both syntactic and semantic information in constraining a parsing decision (Trueswell *et al.*, 1994).

The second issue is the relationship between various syntactic theories. Categorial grammars (Moortgat, 1997; Steedman, 2000), head-driven phrase structure grammar (Pollard and Sag, 1994), lexical-functional grammar (Bresnan, 1982), minimalism (Chomsky, 1995), tree-adjoining grammars (Abeillé and Rambow, 2000), and OT (Legendre *et al.*, 2000) are just some of the syntactic formalisms currently in existence. Significant progress has been made in exploring the connections between their very different representational assumptions (Vijay-Shankar and Weir, 1994; Frank, 2002), and these explorations raise an important issue: if one formalism *A* can be shown to be equivalent, in some well-defined sense, to some other formalism *B*, and if the ontology of *A* is argued to be associated with processing difficulty via some linking assumption to sentence parsing, then this implies that in reality the theoretical commitment of grammar-based approaches is to some common base of the equivalent frameworks rather than to the specific primitives of a particular one. An important step in grammar-based approaches is thus identifying exactly what this common theoretical commitment is – what must be invariant about the grammar to derive the same empirical predictions. For example, is there an empirical motivation for characterizing dislocations such as wh-movement as movement rather than some kind of feature sharing? There have been attempts to distinguish between positing movement-based traces and lexical association accounts (Gibson and Hickok, 1993; Nakano *et al.*, 2000), but much more work remains to be done in this area.

Symbolic Models Involving Complexity Metrics and Ambiguity Resolution Principles

A second kind of model proposes complexity metrics or ambiguity resolution principles predicated on symbolic syntactic structures. The metrics and ambiguity principles can be defined without reference to a particular architecture for the competence grammar, though particular syntactic representations must be assumed in order to make specific predictions. The common underlying assumption of these models is that the metrics or principles are abstract characterizations of the resource-boundedness of human working memory for linguistic processing.

Ambiguity Resolution Principles

The best-known theory of this type is minimal attachment (Frazier, 1979). Minimal attachment is assumed to be a principle that accounts for attachment preferences in ambiguous sentences. Sentences such as *The girl saw the boy with the telescope* are ambiguous because, in the absence of a context, we do not know whether the prepositional phrase *with the telescope* modifies the verb *saw* or the noun *boy*. Minimal attachment provides a simple formal statement of which attachment should be preferred: it is the attachment that yields the simplest structure, where 'simplest' can be defined as the structure that introduces the fewest new syntactic nodes. Thus, minimal attachment depends crucially on the representational assumptions in some syntactic theory. For example, in the prepositional phrase attachment example, the minimal attachment prediction depends on the precise syntactic structure assigned to the VP-adjunct vs. NP-adjunct. Frazier (1979) assumed that the NP-adjunction introduces an additional syntactic node, predicting initial VP attachment in an NP-VP-PP ambiguity.

Another issue involving ambiguity is early vs. late closure (Frazier, 1979). The relevant phenomena can be illustrated in the following way. In a sentence such as *After the student moved the chair broke*, a prosodic break after *moved* (or a comma if the sentence is in written form) will 'close' the phrase *After the student moved* early; this is called early closure, and is the intended utterance. By contrast, if the sentence is spoken (or written) without a prosodic break (or comma) after *moved*, the perceiver tends to 'close' the phrase late, at the end of *the chair*, in this case incorrectly. This initial misanalysis is called late closure. In the absence of any prosodic information (and without any punctuation), such sentences often cause garden-pathing (initial misanalysis due to ambiguity).

Frazier proposed these principles with the specific idea of a constrained-capacity working memory system. Regarding late closure, Frazier (1979: 39) said, "It is a well-attested fact about human memory that the more structured the material to be remembered, the less burden the material will place on immediate memory. Hence, by allowing incoming material to be structured immediately, Late Closure has the effect of reducing the parser's memory load." Similarly, regarding minimal attachment (1979: 40): "The Minimal Attachment strategy not only guarantees minimal structure to be held in memory, but also minimizes rule accessing. Hence, [minimal attachment is also an 'economical' strategy] in the sense that [it reduces] the computation and memory load of the parser." Thus, the limited capacity of working memory was a central motivating factor in the development of theories of ambiguity in sentence processing.

Going beyond the purely capacity-based explanation for ambiguity is Frazier and Clifton's theory of construal (Frazier and Clifton, 1996). One of the goals of this theory is to account for attachment ambiguities like *The girl with the hat that looked funny. . . .* Here, it is unclear who was funny, the girl or the hat. Construal predicts a preference for lower attachment (modification of *hat* rather than *girl)* as follows. Two kinds of relation are assumed, primary and secondary. Primary relations are, for example, those between verbs and their arguments, while secondary relations are relations between modifiers and the element modified, such as the modifying relative clause *that was funny* and the modifiable elements *girl* and *hat*. Frazier and Clifton proposed a construal principle, which (in a simplified form) states that if a nonprimary relation needs to be attached in a syntactic structure, this nonprimary relation must be associated with the phrase containing the most recent theta assigner. This last theta assigner is *with* and the phrase it occurs in is *with the hat*. This accounts for the low attachment preference.

General Structural Complexity Metrics

The theories discussed above focus on ambiguity resolution. Although ambiguity is an extremely important problem for efficient parsing, another important issue is the complexity of globally unambiguous sentences. Several theories attempt to address this issue. The most prominent ones are dependency locality theory (Gibson, 1998, 2000), and early immediate constituents (Hawkins, 1994, 1998, 2001); we describe the first one briefly.

Dependency locality theory (DLT) aims to account for processing difficulty in both ambiguous (garden path) structures and unambiguous ones, such as center embeddings. DLT assumes that during the course of sentence parsing, computational resources in working memory are needed for two aspects: storage of the structure built up thus far and integration of the current word into the structure built up thus far. Based on these two components, a cost metric is defined which predicts relative processing difficulty. There is a storage cost, measured in memory units

Input words:	The	reporter	disliked	the	boy
Heads needed:	Noun, Verb	Verb	NP	Noun	–
Storage cost (MUs):	2	1	1	1	0

Figure 1 Illustration of DLT's storage cost computation.

(MUs), and an integration cost, measured in energy units (EUs).

Storage cost is computed as follows: 1 MU is associated with each syntactic head required to complete the current input as a grammatical sentence. For example, as shown in **Figure 1**, for the first word, *The*, a noun and a verb are needed to complete the sentence grammatically, so the storage cost is 2 MUs. At the second word, *reporter*, only a verb is needed, so the storage cost is 1 MU. For the next word, *disliked*, one NP is needed, so the cost is 1 MU. The cost for the next word, *the*, is also 1 MU, since only a noun is needed to complete the input string as a grammatical sentence. The last word incurs no cost since it can complete the sentence.

The cost of integrating a current word *A* with an existing structure *S* depends on locality, i.e., the distance between them. The processing cost depends on the complexity of the computations that took place between *A* and *S* and is assumed to be linearly related to the number of discourse referents (DRs) introduced between the two items. A DR is assumed to be "an entity that has a spatiotemporal location so that it can later be referred to with an anaphoric expression, such as a pronoun for NPs, or tense on a verb for events" (Gibson, 2000: 103).

Apart from storage cost and integration cost, the other factors that are assumed to affect comprehension difficulty are:

1. The frequency of the lexical item being integrated: the lower the frequency, the greater the processing difficulty.
2. The contextual plausibility of the resulting structure: the less plausible the final structure, the greater the processing difficulty.
3. The discourse complexity of the final structure: nonfocused entities and elements introduced using definite descriptions are less accessible (and are therefore harder to process) than focused entities. Least accessible of all are elements introduced using indefinite NPs: "Elements new to the discourse, which are usually introduced using indefinite NPs, require the most resources because they must be constructed in the discourse model . . . it is assumed that processing the head noun of an NP that refers to a new discourse object consumes substantial resources, and processing the head verb of a VP that refers to a new discourse event (also a discourse referent) consumes substantial

resources; but processing other words does not consume substantial resources in the discourse processing component of structure building" (Gibson, 2000: 103). Also see Warren (2001: 33–44).

4. If the current word is not compatible with the highest ranked structure built so far, there is reanalysis difficulty.

The storage and integration costs together provide a measure of integration time, which gives the complexity at any given point, as well the as overall complexity, of a sentence. Gibson assumed the following regarding these two costs: (i) integrations and storage access the same pool of resources; (ii) resources have fixed capacity; (iii) each predicted syntactic head takes up a fixed quantity of resources; and (iv) the overall acceptability of a sentence depends on the maximal integration time complexity experienced during the parsing process.

In sum, sentence processing according to the DLT is constrained by limited resources in working memory. This limit on resources is quantified in terms of integration and storage costs of linguistic elements as they are processed in real time.

Cognitive Architectures and Sentence Processing

Another approach to sentence-processing theory is to apply independently motivated theories and principles from cognitive psychology research. One instance of such an approach takes as its starting point a fixed cognitive architecture. These theories derive processing explanations from general cognitive processing principles such as activation decay and memory interference.

Activation-Based Architectures

Activation models have dominated the cognitive-architecture-based approaches. In activation-based models, symbolic long- and/or short-term memory elements have a continuous quantity (activation) associated with them that affects processing in some way. Early examples include the Just and Carpenter READER model, which was based on a production system architecture developed concomitantly with the sentence-processing model (Just and Carpenter, 1987). The READER model later evolved into a model of individual differences in language comprehension, with a quantitative parameter (amount of available activation) used to model differences in working memory capacity (Just and Carpenter, 1992). Several other sentence-processing models have been developed in general activation-based architectures developed jointly with a specific language model (Kintsch, 1988; Stevenson, 1994; Vosse and Kempen, 2000; Tabor and Hutchins, 2004; Tabor *et al.*, 2004).

Independent Cognitive Architectures

More recently, there have been attempts to derive sentence-processing models from independent cognitive architectures developed largely outside the domain of linguistic processing. Examples are the SOAR-based architecture in Lewis (1993) and the authors' ACT-R-based sentence-processing model (Lewis and Vasishth, forthcoming). As an example of this approach, and the cognitive architecture approach more generally, we briefly discuss our own model.

Brief Overview of ACT-R

The cognitive theory ACT-R (Anderson *et al.*, 2004) is implemented as a general computational model and incorporates constraints developed through considerable experimental research on human information processing. The ACT-R theory relevant for the present discussion and the parsing model are outlined next, and its empirical coverage is briefly discussed.

In its essence, ACT-R consists of two distinct systems, declarative memory and procedural memory. Declarative memory consists of items (chunks) identified by a single symbol. Each chunk is a set of feature/value pairs; the value of a feature may be a primitive symbol or the identifier of another chunk, in which case the feature/value pair represents a relation.

In addition to the memory systems, focused buffers hold single chunks. There is an architecturally fixed set of buffers, each of which holds a single chunk in a distinguished state that makes it available for processing. Items outside of the buffers must be retrieved to be processed. The three important cognitive buffers are a goal buffer, a problem state buffer, and a retrieval buffer. The goal buffer serves to represent current control state information, and the problem state buffer represents the current problem state. The retrieval buffer serves as the interface to declarative memory, holding the single chunk from the last retrieval. This structure has much in common with conceptions of working memory and short-term memory that posit an extremely limited focus of attention of one to three items, with retrieval processes required to bring items into the focus for processing (Wickelgren *et al.*, 1980; McElree and Dosher, 1993).

All procedural knowledge is represented as production rules (Newell, 1973) – asymmetric associations specifying conditions and actions. Conditions are patterns to match against buffer contents, and actions are taken on buffer contents. All behavior arises from production rule firing; the order of behavior is not fixed in advance but emerges in response to the dynamically changing contents of the buffers.

The sentence-processing model critically depends on the built-in constraints on activation fluctuation of chunks as a function of usage and delay. Chunks have numeric activation values that fluctuate over time; activation reflects usage history and time-based decay. The activation affects a chunk's probability and latency of retrieval.

ACT-R also assumes that associative retrieval is subject to interference. Chunks are retrieved by a content-addressed (McElree, 2000), associative retrieval process. Similarity-based retrieval interference arises as a function of retrieval cue overlap: the effectiveness of a cue is reduced as the number of items associated with the cue increases. Associative retrieval interference arises because the strength of association from a cue is reduced as a function of the number of items associated with the cue.

A Sentence-Processing Model Based on ACT-R

The sentence-processing model consists of a definition of lexical items in permanent memory defined in terms of feature/value pairs, and a set of production rules specifying a left-corner parser (Aho and Ullman, 1977) . The most important property of a left-corner parser in this context is that structure is built as input comes in; there is no waiting period (lookahead; cf. Marcus, 1980) before any parsing decisions are made. The motivation for such an incremental parsing model comes from previous work in the literature (see, for example, Tyler and Marslen-Wilson, 1977; Johnson-Laird, 1983; Scheepers *et al.*, 1999).

The model has been applied to five different published reading experiments (from three laboratories and two different paradigms). The simulations provide detailed accounts of the effects of length and structural interference on both unambiguous and garden path structures. The phenomena covered include an interesting novel prediction of the model, viz., a crossover interaction of the effect of length and structural interference on reanalysis cost vs. initial attachment cost: interference affects attachment more than reanalysis, while the reverse is true for length. The reason for this interaction is that the dispreferred structure does not receive additional activation boosts from participating in the ongoing parse, and so length disproportionately affects its retrieval during garden path reanalysis. This retrieval does not, however, incur greater interference cost, because that cost is determined by the intervening structures, which are often the same regardless of which interpretation was pursued. Van Dyke and Lewis (2003) obtained quantitative estimates of these effects and their interaction, using ambiguous and unambiguous versions of the constructions in (1). The model provides accounts of reading times in the critical disambiguating region and the ambiguous region (the latter testing predictions of load effects).

(1a) [short] The assistant forgot (that) the student was standing in the hallway.

(1b) [long/simple] The assistant forgot (that) the student who was waiting for the exam was standing in the hallway.

(1c) [long/complex] The assistant forgot (that) the student who knew that the exam was important was standing in the hallway.

An important result in this model is that all fits were obtained with one free scaling parameter that was fixed across all the simulations; all remaining quantitative parameters were set to default values from the ACT-R literature. The remaining theoretical degrees of freedom in the model are the production rules that embody the parsing skill, and these rules represent a straightforward realization of left-corner parsing conforming to one overriding principle: compute the parse as quickly as possible. This approach thereby considerably reduces theoretical degrees of freedom – both the specific nature of the strategic parsing skill and the mathematical details of the memory retrieval derive from existing theory, plus the assumption of fast, incremental parsing.

In summary, this sentence-processing model (a) provides an integrated, quantitative account of both length and structural complexity effects in both ambiguous and unambiguous constructions; (b) predicts reading times in both ambiguous and disambiguating regions; (c) probabilistically predicts both parsing failures and reading times; and (d) provides single-parameter quantitative predictions across multiple experiments and paradigms.

Probabilistic Models

In probabilistic models, the starting point is a set of symbolic rules that generate syntactic structures (an example is a context-free grammar specification). To such rules a probability is assigned in the following way: the probability of a rule's left-hand side expanding to the right-hand side is assigned a numerical value between 0 and 1, this number being the product of the probabilities of the right-hand side nonterminal symbols. All rules with the same left-hand side have a total probability summing to one, and a parse tree's probability is a product of the set of all rules used to generate the tree (Manning and Schütze, 1999).

Ambiguity Resolution Using a Probabilistic Approach

Probabilistic models capture an important aspect of sentence processing: the role of experience and frequency in language processing. They also furnish good models of parsing preferences, of ambiguity resolution, and of gradedness in acceptability (Crocker and Keller, forthcoming).

We present next an instance of a probabilistic approach and show how it applies to a well-known ambiguity resolution issue (Jurafsky, 1996). This model takes as its starting point a syntactic formalism known as construction grammar. Construction grammar assumes that complex, nonlocal constructions are stored along with lexical entries in the mental grammar. Jurafsky's model assigns probabilities to each such construction, and these probabilities determine the likelihood of their being accessed. The probabilities, which are analogous to initial activation values of entities in the mental grammar, are assigned using a corpus (the Penn Treebank) to compute maximum likelihood estimates from relative frequencies. In addition, valency frequencies of verbs are taken from Connine *et al.* (1984).

The classic garden path sentence, *the horse raced past the barn fell*, is then explained as follows. Of the two possible interpretations, one has *raced* as an intransitive verb (main verb interpretation) and the other has *raced* as a transitive (reduced relative interpretation). The intransitive valence has a probability of 0.92, while the transitive one has a probability of 0.08. In addition the reduced relative interpretation requires a context-free grammar (CFG) rule to be applied as in (2):

(2) $NP \rightarrow NP\ XP$

This rule's probability (computed from the Penn Treebank) is 0.14. The combination of the transitive verb valency probability and the CFG rule is $0.14 \times 0.08 = 0.0112$. Since this combined probability of the reduced relative interpretation is $0.92 \div 0.0112 = 82.14286$ times lower than that of the preferred interpretation, it is pruned away during processing because of an assumed beam search mechanism that rejects any parses that are too far away from the best.

Some Open Issues in Probabilistic Models

Two important issues in sentence processing that remain to be addressed in detail are: explaining moment-by-moment fluctuations in processing difficulty, and the relative difficulty of complex yet unambiguous sentences. The ability to explain reading time data gathered by methods like self-paced reading and eyetracking requires further refinement of probabilistic models, which generally provide a single real number for the full sentence parse, which serves as a measure of processing difficulty. The second issue is that current probabilistic models do not have as broad an account as some other approaches (Gibson, 2000; Hawkins, 2001; Lewis and Vasishth, forthcoming) of globally unambiguous sentences, such as subject vs. object relative clause processing, center embedding phenomena, etc. Some progress has been made in this direction (Crocker and Brants, 2000; Hale, 2001; Korthals, 2001; Narayanan and Jurafsky, 2002, Jurafsky, 2003; Vasishth and Uszkoreit, 2004), but much more remains to be done.

Symbolic Processing Models and Their Relation to Connectionist Models

We conclude by briefly considering symbolic models in terms of several properties normally associated with connectionist models: parallel processing, interactivity (simultaneous application of multiple constraints), and the graded, continuous nature of human performance.

First, it should be clear that parallel vs. serial processing is a distinction that is orthogonal to symbolic vs. nonsymbolic processing. For example, all of the prominent symbolic cognitive architectures (Soar [Newell, 1990], EPIC [Kieras and Meyer, 1997], and ACT-R [Anderson *et al.*, 2004]) include substantial amounts of parallelism. In the domain of sentence processing, most symbolic models also assume parallel processes – for example, the ACT-R-based model above assumes that memory retrieval happens in parallel with other cognitive and perceptual processes.

It is also worth clarifying here a potential source of confusion in the sentence processing literature that relates to the serial vs. parallel processing issue (Lewis, 2000b). A distinction is often made between serial and parallel parsing, but these terms are generally meant to refer to depth-first search vs. breadth-first search methods, respectively – that is, serial parsing pursues a single interpretation while parallel parsing pursues multiple interpretations. But this distinction is also orthogonal to serial vs. parallel processing as it is used elsewhere in the connectionist and symbolic modeling literature, where it simply refers to the temporal sequentiality or overlap of processes. Thus, one can have a serial-processing parallel parser, or a parallel-processing serial parser. In fact, the ACT-R model described above is an example of the latter.

Second (and related to parallel processing), simultaneous constraint interaction and constraint satisfaction from different sources (e.g., the immediate effect of context on parsing) is also completely consistent with symbolic processing. One example is the modular symbolic architecture sketched in Altmann and Steedman (1988), which involves multiple constraint satisfaction from different sources.

Finally, although we defined symbolic-processing systems as ones that manipulate relatively discrete combinatorial representations, such systems are still capable of accounting for aspects of the continuous, graded nature of human performance. Examples are the probabilistic approaches in Hale (2004) and Crocker and Keller (forthcoming), optimality theoretic approaches, and the ACT-R model mentioned in this article (Lewis and Vasishth, forthcoming).

Looking ahead, it seems clear that labels such as 'symbolic' and 'connectionist' will continue to be at best vaguely descriptive categories, and sharper theoretical discussion will appeal to precisely articulated issues such as the nature of the parsing search strategy, the precise relationship of parallel processes in terms of memory and control, the distribution of linguistic knowledge across the architectural components, the nature of codified experience, and the interaction of language-specific and more general cognitive resources. The current class of symbolic cognitive models comprises a rich body of theory from which to pursue these and other pressing research questions.

See also: Generative Grammar; Human Language Processing: Connectionist Models; Information Theory; Language: Mathematical Complexity; Language Processing: Statistical Methods; Parsing: Statistical Methods; Psycholinguistics: Overview; Rational Analysis and Language Processing; Sentence Processing; Syntax: Optimality Theory.

Bibliography

Abeillé A & Rambow O (eds.) (2000). *Tree adjoining grammars.* Stanford, CA: CSLI.

Aho A V & Ullman J D (eds.) (1977). *Principles of compiler design.* Reading, MA: Addison-Wesley.

Altmann G & Steedman M (1988). 'Interaction with context during human sentence processing.' *Cognition 30,* 191–238.

Anderson J, Bothell D, Byrne M & Lebiere C (2004). 'An integrated theory of the mind.' *Psychological Review 111(4),* 1036–1060.

Bach E, Brown C & Marslen-Wilson W (1986). 'Crossed and nested dependencies in German and Dutch: a psycholinguistic study.' *Language and Cognitive Processes 1(4),* 249–262.

Berwick R C & Weiriberg A S (eds.) (1984). *The grammatical basis of linguistic performance.* Cambridge, MA: MIT Press.

Bresnan J (1978). 'A realistic transformational grammar.' In Halle M, Bresnan J & Miller G (eds.) *Linguistic theory and psychological reality.* Cambridge, MA: MIT Press. 1–59.

Bresnan J (1982). *The mental representation of grammatical relations.* Cambridge, MA: MIT Press.

Bresnan J & Kaplan R (1982). 'Introduction: grammars as mental representations of language.' In Bresnan J (ed.) *The mental representation of grammatical relations.* Cambridge, MA: MIT Press. xvii–lii.

Carpenter P A, Miyake A & Just M A (1995). 'Language comprehension: sentence and discourse processing.' *Annual Review of Psychology 46,* 91–120.

Chomsky N (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky N (1977). 'On wh-movement.' In Culicover P, Wasow T & Akmajian A (eds.) *Formal syntax.* New York: Academic Press. 71–132.

Chomsky N (1995). *The minimalist program.* Cambridge, MA: MIT Press.

Connine C, Ferreira F, Jones C & Abney S (1984). 'Verb frame preference: descriptive norms.' *Journal of Psycholinguistic Research 13,* 307–319.

Crocker M W & Brants T (2000). 'Wide-coverage probabilistic sentence processing.' *Journal of Psycholinguistic Research 29(6),* 647–669.

Crocker M & Keller F (2005). 'Probabilistic grammars as models of gradience in language processing.' In Fanselow G, Féry C, Vogel R & Schlesewsky M (eds.) *Gradience in grammar: generative perspectives.* Oxford University Press.

Fanselow G, Schlesewsky M, Cavar D & Kliegl R (1999). *Optimal parsing, syntactic parsing preferences, and optimality theory.* New Brunswick: Rutgers Optimality Archive, Rutgers State University of New Jersey.

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

Frank R (2002). *Phrase structure composition and syntactic dependencies.* Cambridge, MA: MIT Press.

Frazier L (1979). On comprehending sentences: syntactic parsing strategies. Unpublished dissertation, University of Massachusetts, Amherst.

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

Gibson E (1998). 'Linguistic complexity: locality of syntactic dependencies.' *Cognition 68,* 1–76.

Gibson E (2000). 'Dependency locality theory: a distance-based theory of linguistic complexity.' In Marantz A, Miyashita Y & O'Neil W (eds.) *Image, language, brain: papers from the first mind articulation project symposium.* Cambridge, MA: MIT Press.

Gibson E & Hickok G (1993). 'Sentence processing with empty categories.' *Language and Cognitive Processes 8,* 147–161.

Hale J (2001). 'A probabilistic Earley parser as a psycholinguistic model.' In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics,* Pittsburgh, PA.

Hale J (2004). 'The information-processing difficulty of incremental parsing.' In *Proceedings of the workshop on Incremental Parsing, Association of Computational Linguistics,* Barcelona.

Hale J & Smolensky P (2001). *A parser for harmonic context-free grammars.* Baltimore: Johns Hopkins University.

Hawkins J A (1994). *A performance theory of order and constituency.* New York: Cambridge University Press.

Hawkins J A (1998). 'Some issues in a performance theory of word order.' In Siewierska A (ed.) *Constituent order in the languages of Europe.* Berlin: Mouton de Gruyter.

Hawkins J A (2001). 'Why are categories adjacent?' *Journal of Linguistics 37*, 1–34.

Hoeks J & Hendriks P (2004). Optimality theory and human sentence processing: towards a cross-modular analysis of coordination. Unpublished MS, University of Groningen.

Johnson-Laird P N (1983). *Mental models: towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press.

Joshi A K (1990). 'Processing crossed and nested dependencies: an automaton perspective on the psycholinguistic results.' *Language and Cognitive Processes 5(1)*, 1–27.

Jurafsky D (1996). 'A probabilistic model of lexical and syntactic access and disambiguation.' *Cognition 20*, 137–194.

Jurafsky D (2003). 'Probabilistic modeling in psycholinguistics: linguistic comprehension and production.' In Bod R, Hay J & Jannedy S (eds.) *Probabilistic linguistics.* Cambridge, MA: MIT Press.

Just M A & Carpenter P A (1987). *The psychology of reading and language comprehension.* Boston: Allyn and Bacon.

Just M A & Carpenter P A (1992). 'A capacity theory of comprehension: individual differences in working memory.' *Psychological Review 99(1)*, 122–149.

Kayne R S (1994). *The antisymmetry of syntax.* Cambridge, MA: MIT Press.

Kieras D E & Meyer D E (1997). 'An overview of the EPIC architecture for cognition and performance with application to human–computer interaction.' *Human–Computer Interaction 12*, 391–438.

Kintsch W (1988). *Comprehension: a paradigm for cognition.* New York: Cambridge University Press.

Korthals C (2001). 'Self embedded relative clauses in a corpus of German newspaper texts.' In Striegnitz K (ed.) *Proceedings of the Sixth ESSLLI Student Session.* Helsinki: University of Helsinki. 179–190.

Legendre G, Grimshaw J & Vikner S (2000). *Optimality-theoretic syntax.* Cambridge, MA: MIT Press.

Lewis R L (1993). An architecturally-based theory of human sentence comprehension. Unpublished dissertation, Carnegie Mellon University.

Lewis R L (2000a). 'Cognitive modeling, symbolic.' In Wilson R & Keil F (eds.) *The MIT encyclopedia of the cognitive sciences.* Cambridge, MA: MIT Press.

Lewis R L (2000b). 'Falsifying serial and parallel parsing models: empirical conundrums and an overlooked paradigm.' *Journal of Psycholinguistic Research 29(2)*, 241–248.

Lewis R L & Vasishth S (2005). 'An activation-based model of sentence processing as skilled memory retrieval.' *Cognitive Science 29*, 375–419.

Manning C & Schütze H (1999). *Foundations of statistical natural language processing.* Cambridge, MA: MIT Press.

Marcus M P (1980). *A theory of syntactic recognition for natural language.* Cambridge, MA: MIT Press.

McElree B (2000). 'Sentence comprehension is mediated bycontent-addressable memory structures.' *Journal of Psycholinguistic Research 29(2)*, 111–123.

McElree B & Dosher B A (1993). 'Serial retrieval processes in the recovery of order information.' *Journal of Experimental Psychology: General 122(3)*, 291–315.

Miller G & Chomsky N (1963). 'Finitary models of language users.' In Luce R D, Bush R R & Galanter E (eds.) *Handbook of mathematical psychology,* 3 vols. New York: John Wiley. Vol. 2, 419–492.

Moortgat M (1997). 'Categorial type logics.' In Van Benthem J & Ter Meulen A (eds.) *Handbook of logic and language.* Amsterdam: Elsevier.

Nakano Y, Fesler C & Clahsen H (2000). Antecedent priming at trace positions in Japanese long-distance scrambling. MS, University of Essex.

Narayanan S & Jurafsky D (2002). 'Combining structure and probabilities in a Bayesian model of human sentence processing.' In *Proceedings of the 15th CUNY Sentence Processing Conference.* New York.

Newell A (1973). 'Production systems: models of control structures.' In Chase W G (ed.) *Visual information processing.* New York: Academic Press.

Newell A (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Phillips C (1990). *Order and structure.* Unpublished dissertation, MIT.

Pollard C & Sag I A (1994). *Head-driven phrase structure grammar.* Chicago: University of Chicago Press.

Rambow O & Joshi A K (1994). 'A processing model for free word order languages.' In Clifton J C, Frazier C L & Rayner K (eds.) *Perspectives on sentence processing.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Scheepers C, Hemforth B & Konieczny L (1999). 'Incremental processing of German verb-final constructions: predicting the verb's minimum (!) valency.' In *Proceedings of the Second International Conference on Cognitive Science.* Tokyo.

Shannon C E (1948). 'A mathematical theory of communication.' *Bell System Technical Journal 27*, 379–423.

Slobin D I (1966). 'Grammatical transformations and sentence comprehension in childhood and adulthood.' *Journal of Verbal Learning and Verbal Behavior 5*, 219–227.

Smolensky P & Legendre G (2005). *The harmonic mind: from neural computation to optimality.* Cambridge, MA: MIT Press.

Steedman M (2000). *The syntactic process.* Cambridge, MA: MIT Press.

Stevenson S (1994). A competitive attachment model for resolving syntactic ambiguities in natural language parsing. Unpublished dissertation, University of Maryland.

Stevenson S & Smolensky P (2005). 'Optimality in sentence processing.' In Smolensky P & Legendre G (eds.) *The harmonic mind: from neural computation to optimality-theoretic grammar. Volume II. Linguistics and philosophical implications.* Cambridge, MA: MIT Press.

Tabor W & Hutchins S (2004). 'Evidence for self-organized sentence processing: digging-in effects.' *Journal of Experimental Psychology: Learning, Memory, and Cognition 30(2)*, 431–450.

Tabor W, Galantucci B & Richardson D (2004). 'Effects of merely local syntactic coherence on sentence processing.' *Journal of Memory and Language 50*, 355–370.

Trueswell J C, Tanenhaus M & Garnsey S M (1994). 'Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution.' *Journal of Memory and Language 33*, 285–318.

Tyler L K & Marslen-Wilson W D (1977). 'The on-line effects of semantic context on syntactic processing.' *Journal of Verbal Learning and Verbal Behavior 16*, 683–692.

Van Dyke J & Lewis R L (2003). 'Distinguishing effects of structure and decay on attachment and repair: a cue-based parsing account of recovery from misanalyzed ambiguities.' *Journal of Memory and Language 49*, 285–316.

Vasishth S & Uszkoreit H (2004). 'Self center embeddings revisited.' In *Proceedings of the Architectures and Mechanisms for Language Processing Conference.* France: Aix-en-Provence.

Vijay-Shankar K & Weir D (1994). 'The equivalence of four extensions of context-free grammars.' *Mathematical Systems Theory 27*, 511–545.

Vosse T & Kempen G (2000). 'Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and lexicalist grammar.' *Cognition 75*, 105–143.

Warren T C (2001). Understanding the role of referential processing in sentence complexity. Unpublished dissertation, MIT.

Wickelgren W A, Corbett A T & Dosher B A (1980). 'Priming and retrieval from short-term memory: a speed–accuracy tradeoff analysis.' *Journal of Verbal Learning and Verbal Behavior 19*, 387–404.

Human Language Technology

G Hirst, University of Toronto, Toronto, Canada

Human language technology, or *technologies*, or *HLT*, is the name given collectively to the various areas of technology, especially computer technology, that involve applications to tasks in which language is central. Traditionally, three subfields have been recognized: speech technology, information or document retrieval, and natural language processing. Each is described in detail in a separate article in this encyclopedia, as are many of the specific topics of research and areas of application within each subfield. The use of the name *human language technology*, however, is usually intended to emphasize the unity of the field, the integration of elements from different subfields, and an emphasis on application rather than theory. For example, all three subfields would be involved in creating a system that took as its input video or audio recordings of news broadcasts, and automatically transcribed the soundtrack (using speech recognition) to create a written summary of the broadcast (using methods from natural language processing) and to retrieve segments of the video or audio in response to topical requests from a user (using methods from information retrieval).

The three subfields arose as distinct enterprises in separate disciplines with differing research methods and traditions – natural language processing in computer science, speech technology in electrical engineering and digital signal processing, and document and information retrieval in library science – and they have only recently been seen as together forming an integrated area of research with significant interests in common. Early moves in this direction were initiated by U.S. agencies that funded research in the subfields: the first Human Language Technology Conference was held in 1993 with the aim of bringing agency-funded researchers together to learn one another's methods, goals, and interests. In 2001, the conferences became an annual event and were opened to the general international research community; in 2003, their organization was turned over to the Association for Computational Linguistics (*see* **Association for Computational Linguistics**).

A related term, often used in Europe particularly since the early 1990s, is *language engineering* or *natural language engineering* (the latter being also the name of a research journal). While perhaps in practice natural language engineering is more centered on textual applications than on speech, this term, too, implies an orientation toward applications involving language with less regard for the discipline or subfield in which the work arises. Both terms are now commonly used in Europe.

See also: Association for Computational Linguistics; Document Retrieval, Automatic; Natural Language Processing: Overview; Symbolic Computational Linguistics: Overview.

Relevant Websites

www.aclweb.org. – Association for Computational Linguistics.

journals.cambridge.org. – *Natural Language Engineering.*