

Sentence processing

Roger P.G. van Gompel

Department of Psychology

University of Dundee

Dundee DD1 4HN

Scotland, United Kingdom

E-mail: r.p.g.vangompel@dundee.ac.uk

In K. Brown et al. (Eds.) (2006), *Encyclopedia of Language and Linguistics* (2nd edition).

Oxford: Elsevier.

Abstract

Sentence processing research investigates how people process the syntactic structure of sentences, with a particular interest in syntactic ambiguity. Two largely incompatible classes of theories dominate research on syntactic ambiguity resolution. According to two-stage theories, the sentence processor initially adopts a single analysis using only a restricted range of information. In contrast, constraint-based theories claim that multiple analyses of a syntactic ambiguity are activated in parallel and that the processor immediately uses all sources of information. This article reviews experimental studies that provide evidence for and against both theories. It also outlines current controversies in the field and describes new directions that sentence research has taken.

An essential part of understanding a sentence is to construct the appropriate syntactic structure. Research investigating how language users do this is usually referred to as sentence processing research. Most sentence processing research has focused on syntactic ambiguities. Although language users are often unaware of syntactic ambiguities, they are in fact very common, so an important goal of this research is to develop theories that explain how language users process them. Furthermore, syntactic ambiguities present a precious opportunity to gain insight into the architecture and mechanisms of the sentence processor. People often experience difficulty in adopting the intended analysis of syntactically ambiguous sentences and by examining when processing difficulty occurs, psycholinguists can study the workings of the sentence processor.

Sentence processing theories: Two-stage versus constraint-based theories

The literature on syntactic ambiguity resolution is dominated by two largely incompatible classes of theories, two-stage and constraint-based theories. Two-stage theories claim that the processor initially draws upon a restricted range of information to construct a single analysis of an ambiguous structure. According to the most influential two-stage account, the garden-path theory, the syntactic processor is modular (e.g., Frazier, 1979, 1987; Rayner, Carlson, & Frazier, 1983). It claims that the processor initially adopts the analysis that is simplest in terms of syntactic tree structure (e.g., the fewest nodes), whereas other potentially useful information for resolving syntactic ambiguities is delayed until the second stage of processing. Other two-stage theories assume that the processor uses thematic role information during its initial stage (e.g., Pritchett, 1992). However, all two-stage theories claim that reanalysis occurs when the initial analysis is inconsistent with information that is used during the second stage of processing. Discovery of the misanalysis and subsequent reanalysis result in processing difficulty.

In contrast, constraint-based theories claim that all sources of information, including non-syntactic constraints, immediately affect syntactic ambiguity resolution (e.g., MacDonald, Seidenberg, & Pearlmuter, 1994; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993). They claim that the processor activates multiple analyses of a syntactic ambiguity in parallel. All sources of information have an immediate effect on their activation and the activation of an analysis is determined by the number and strength of the constraints supporting it. When the activation of one analysis is much higher than that of the alternative analyses, processing the ambiguity is easy. But when two analyses are approximately equally activated, they compete, and processing difficulty occurs.

The use of non-syntactic information

A large body of research has tested the two opposing classes of theories by investigating whether the use of non-syntactic information is delayed relative to syntactic information. Ferreira and Clifton (1986) investigated sentences such as (1-4) in a reading experiment.

1. The defendant examined by the lawyer turned out to be unreliable.
2. The defendant that was examined by the lawyer turned out to be unreliable.
3. The evidence examined by the lawyer turned out to be unreliable.
4. The evidence that was examined by the lawyer turned out to be unreliable.

Sentence (1) contains a reduced relative clause (*examined by the lawyer*) that results in a temporary ambiguity, because the initial words also permit a main clause analysis (as in *the defendant examined the lawyer*). According to the garden-path theory, the processor initially adopts the main clause analysis, because it is structurally simplest. This correctly accounts for

Ferreira and Clifton's (1986) finding that the disambiguating phrase (*by the lawyer*) in (1) is harder to read than in an unambiguous sentence like (2).

More interestingly, Ferreira and Clifton (1986) observed that readers experienced difficulty with reduced relatives even in (3), where the main clause analysis is ruled out by plausibility (evidence does not normally examine anything). This suggests that the processor bases its initial analysis on syntactic information and that the use of plausibility information is delayed, as claimed by the garden-path theory. However, Trueswell et al. (1994) argued that the plausibility manipulation in Ferreira and Clifton's study was too weak. With stronger materials, they showed that no difficulty occurred in (3) relative to (4). They argued that this supports constraint-based theories.

Other studies have investigated the use of contextual information. Constraint-based theories incorporate ideas from referential theory (Altmann & Steedman, 1988; Crain and Steedman, 1985), which claims that the processor favors the analysis with the fewest unsupported discourse presuppositions. In (5), the temporarily ambiguous PP *with the new lock* modifies the NP *the safe*.

5. The burglar blew open the safe with the new lock.

This interpretation presupposes that there is more than one safe in the discourse, but only one has a new lock. However, in the absence of a context mentioning more than one safe, this presupposition is unsupported. Hence, (5) should be harder to process than the verb modifier interpretation in (6), which does not contain this presupposition.

6. The burglar blew open the safe with the dynamite.

However, in a context mentioning more than one safe, as in (7), the NP modifier interpretation should be preferred.

7. The burglar saw a safe with a new lock and a safe with an old lock.

He blew open the safe with the new lock.

The presupposition associated with the NP modifier is now supported, because the PP singles out one of the safes. In contrast, the verb modifier interpretation is infelicitous because it presupposes that the discourse contains only a single safe.

Several reading studies have shown that referential context affects syntactic ambiguity resolution (e.g., Altmann & Steedman, 1988), however, the effects are less strong than claimed by referential theory: In syntactic ambiguities with a strong bias for one analysis (in the absence of a context), discourse context does not appear to override structural preferences (e.g., Britt, 1994). This fits well with constraint-based theories. They claim that discourse information has a strong effect when other constraints do not strongly favor a single analysis. But when other constraints strongly favor one analysis, these constraints may override context effects.

Another important non-syntactic constraint is the frequency with which structures occur in a language. Trueswell et al. (1993) investigated temporarily ambiguous sentences such as (8-9).

8. The student forgot the solution was in the book.

9. The student hoped the solution was in the book.

Verbs such as *forget* occur more frequently with a direct object (as in *The student forgot the solution*) than with a sentence complement (as in (8)), whereas verbs such as *hope* show the

opposite pattern. Trueswell et al. showed that the disambiguating region in (8) (*was in the*) took longer to read than in unambiguous sentences containing *that* following the critical verb. However, no difference was observed between (9) and its unambiguous control, suggesting that frequency information had an immediate effect on syntactic ambiguity resolution. They concluded that this supports constraint-based theories.

On-going controversies

The garden-path model may be able to account for early effects of non-syntactic information by stipulating that such information is used so rapidly that its delay is undetectable (Clifton & Ferreira, 1986), but this makes its predictions rather unclear. As a result, constraint-based theories have become the dominant approach in the sentence processing literature. However, not all studies support the constraint-based view. Binder, Duffy, and Rayner (2001) investigated whether non-syntactic information (plausibility and context) could make the main clause analysis in reduced relative/main clause ambiguities the preferred analysis, as predicted by constraint-based theories, but it did not. Clifton, Traxler, Mohamed, Williams, Morris, and Rayner (2003) investigated the use of plausibility information in reduced relative structures, but used a larger set of materials than Trueswell et al. (1994). They observed that plausibility information facilitated the reduced relative analysis, but it did not completely eliminate difficulty. Finally, some studies show that verb frequency information does not always eliminate processing difficulty (e.g., Pickering, Traxler, & Crocker, 2000).

Constraint-based theorists argue that such results do not provide evidence against constraint-based theories. Syntactic preferences may be particularly strong and therefore, non-syntactic information cannot always override them. Unfortunately, such claims make it difficult to derive exact and testable predictions as to how linguistic factors interact. In order to make

constraint-based models more explicit and testable, researchers have started to build computational models (e.g., McRae, Spivey-Knowlton, & Tanenhaus, 1998; Tabor & Tanenhaus, 1999). Furthermore, constraint-based theorists have argued that people's completions of sentence fragments provide a good estimate of usage frequencies, and therefore, they can be used to predict parsing preferences (e.g., McRae et al, 1998; Garnsey et al., 1997). However, it is unclear whether parsing preferences are the result of usage frequencies, or whether the same underlying factors influence both parsing preferences and usage frequencies. Hence, a promising alternative approach is to explain both parsing and production preferences as the result of the same semantic factors (e.g., McKoon & Ratcliff, 2003).

New directions

Recent studies have started to address new issues. An important line of research explores the interactions between visual context and sentence processing by measuring people's eye movements to visual scenes while they listen to sentences. Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002) asked people to follow auditory instructions such as (10-11).

10. Put the apple on the towel in the box.

11. Put the apple that's on the towel in the box.

They were presented with either a one-referent scene containing a single apple on a towel, or a two-referent scene containing two apples, one of which was on a towel. Both scenes also contained an empty towel without an apple, and a box. When people heard (10) while looking at the one-referent scene, they looked more often at the empty towel than when hearing (11).

This indicates that they initially misinterpreted *on the towel* in the temporarily ambiguous sentence (10) as modifying the verb *put* and took it as the destination for the apple. However, for the two-referent scene, no such difference was observed, suggesting that the visual context immediately affected syntactic ambiguity resolution. Recent research has also used this visual-world method to investigate how children process syntactic ambiguities. Trueswell, Sekerina, Hill, and Logrip (1999) showed that, in contrast to adults, children do not use the visual context. However, both adults and children appear to use verb frequency information (Snedeker & Trueswell, 2004).

Using a similar method, Altmann and Kamide (1999; Kamide, Altmann, & Haywood, 2003) showed that listeners looked at edible objects when hearing sentences such as (12), even before they heard *cake*.

12. The boy will eat the cake.

Kamide et al. (2003) argued that the processor anticipates properties of upcoming phrases, using both semantic and syntactic information up to that point. What properties it anticipates is less clear, in particular whether it anticipates syntactic structures. Furthermore, it is likely that the anticipations are at least partly triggered by the visual context. During normal language comprehension, such a context is usually absent, and people may therefore not make the same type of anticipations. Nevertheless, what these studies show is that visual context has a very rapid effect on sentence comprehension.

Other recent research has continued to use reading methods, but has addressed new questions. Van Gompel, Pickering, Pearson, and Liversedge (in press) investigated whether processing difficulty is the result of competition between analyses. Most constraint-based theories claim that competition should occur in globally ambiguous sentences with no strong

bias for either analysis, because constraints support two analyses to an equal extent. Hence, such sentences should be harder to process than disambiguated sentences, where only a single analysis is supported. However, Van Gompel et al. observed exactly the opposite pattern of results, suggesting that competition is not the mechanism causing processing difficulty. They argued that processing difficulty is due to reanalysis: Globally ambiguous sentences are easy to process, because the initial analysis remains possible throughout the sentence, regardless of which analysis is adopted. In contrast, when a sentence is disambiguated, the initial analysis may be incorrect, and the processor has to reanalyze.

Christianson, Hollingworth, Halliwell, and Ferreira (2001) showed that readers often fail to adopt the intended analysis in temporarily ambiguous sentences. In line with this, Ferreira (2003) argued that readers often construct shallow representations of the sentence, even for completely unambiguous sentences such as passives. Finally, researchers have started to show a renewed interest in how working memory constraints affect the processing of both ambiguous and unambiguous sentences (e.g., Caplan & Waters, 1999; Gibson, 1998; Gordon, Hendrick, & Johnson, 2004). Future research will probably continue to explore these and other new issues.

Bibliography

- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, **30**, 191-238.
- Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, **73**, 247-264.
- Binder, K.S., Duffy, S.A., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language*, **44**, 297-324.
- Britt, M.A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, **33**, 251-283.
- Caplan, D., & Waters, G.S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, **22**, 77-126.
- Christianson, K., Hollingworth, A., Halliwell, J.F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, **42**, 368-407.
- Clifton, C.J., & Ferreira, F. (1989). Ambiguity in context. *Language and Cognitive Processes*, **4**, SI 77-103.
- Clifton, C., Traxler, M.J., Mohamed, M.T., Williams, R.S., Morris, R.K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, **49**, 317-334.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D.R. Dowty, L. Karttunen, & A.M. Zwicky (Eds.), *Natural Language Parsing: Psychological, Computational and Theoretical perspectives* (pp. 320-358). Cambridge, England: CUP.
- Ferreira, F., & Clifton, C.J. (1986). The independence of syntactic processing. *Journal of Memory and Language*, **25**, 348-368.

- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, **47**, 164-203.
- Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Ph.D. Dissertation. Indiana University Linguistics Club. University of Connecticut.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559-586). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, **68**, 1-76.
- Gordon, P.C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, **51**, 97-114.
- Kamide, Y., Altmann, G.T.M., & Haywood, S.L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, **49**, 133-156.
- MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676-703.
- McKoon, G., & Ratcliff, R. (2003). Meaning through syntax: Language comprehension and the reduced relative clause construction. *Psychological Review*, **110**, 490-525.
- McRae, K., Spivey-Knowlton, M.J., & Tanenhaus, M.K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, **38**, 283-312.
- Pickering, M.J., Traxler, M.J., & Crocker, M.W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, **43**, 447-475.

- Pritchett, B.L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, **22**, 358-374.
- Snedeker, J., & Trueswell, J.C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, **49**, 238-299.
- Spivey, M.J., Tanenhaus, M.K., Eberhard, K.M., & Sedivy, J.C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, **45**, 447-481.
- Tabor, W., & Tanenhaus, M.K. (1999). Dynamical models of sentence processing. *Cognitive Science*, **23**, 491-515.
- Tanenhaus, M.K., Spivey Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, **268**, 1632-1634.
- Trueswell, J.C., Tanenhaus, M.K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology Learning, Memory, and Cognition*, **19**, 528-553.
- Trueswell, J.C., Tanenhaus, M.K., & 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.
- Trueswell, J.C., Sekerina, I., Hill, N.M., & Logrip, M.L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, **73**, 89-134.

Van Gompel, R.P.G., Pickering, M.J., Pearson, J., & Liversedge, S.P. (in press). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*.