ORIGINAL ARTICLE

In Defense of Competition During Syntactic Ambiguity Resolution

Theo Vosse · Gerard Kempen

Abstract   In a recent series of publications (Traxler et al. J Mem Lang 39:558–592, 1998; Van Gompel et al. J Mem Lang 52:284–307, 2005; see also Van Gompel et al. (In: Kennedy, et al.(eds) Reading as a perceptual process, Oxford, Elsevier pp 621–648, 2000); Van Gompel et al. J Mem Lang 45:225–258, 2001) eye tracking data are reported showing that globally ambiguous (GA) sentences are read faster than locally ambiguous (LA) counterparts. They argue that these data rule out "constraint-based" models where syntactic and conceptual processors operate concurrently and syntactic ambiguity resolution is accomplished by competition. Such models predict the opposite pattern of reading times. However, this argument against competition is valid only in conjunction with two standard assumptions in current constraint-based models of sentence comprehension: (1) that syntactic competitions (e.g., Which is the best attachment site of the incoming constituent?) are pooled together with conceptual competitions (e.g., Which attachment site entails the most plausible meaning?), and (2) that the duration of a competition is a function of the overall (pooled) quality score obtained by each competitor. We argue that it is not necessary to abandon competition as a successful basis for explaining parsing phenomena and that the above-mentioned reading time data can be accounted for by a parallel-interactive model with conceptual and syntactic processors that do not pool their quality scores together. Within the individual linguistic modules, decision-making can very well be competition-based.

Keywords   Syntactic processing · Sentence comprehension · Ambiguity resolution · Competition · Reanalysis

In present-day psycholinguistic theorizing on ambiguity resolution in human sentence comprehension, one often distinguishes two types of parsing strategies: *syntax-first* (or *garden-path*) and *constraint-based* (cf. Van Gompel et al. 2000). Both types of models

T. Vosse
FC Donders Centre for Cognitive Neuroimaging, P.O. Box 9101, Nijmegen 6500 HB, The Netherlands
e-mail: theo.vosse@fcdonders.ru.nl

G. Kempen (✉)
Max Planck Institute for Psycholinguistics, P.O. Box 310, Nijmegen 6500 AH, The Netherlands
e-mail: gerard.kempen@mpi.nl

presuppose that the ultimately delivered *syntactic* structure for a grammatically well-formed sentence has to be compatible with the *conceptual* (semantic/pragmatic) content expressed in the sentence. They are also similar in assuming incremental processing: The syntactic parser attempts to assign a structural place/role to every new input word immediately, without waiting for the rest of the sentence. The two types of model differ with respect to the time course of syntactic and conceptual processing. According to syntax-first models, when encountering an ambiguous input string, the parser derives one syntactic structure during a first stage of processing. Here, only the morphological and syntactic properties of the new input word (in particular its word-class) are taken into account. After thus having computed a syntactic structure for the ambiguous input string up to and including the most recent input word, the parser submits it to a conceptual processor, which attempts to assign it a plausible meaning. If this second processing stage does not yield a satisfactory result, the syntactic parser concludes it has been "led up the garden path" and undertakes a reanalysis of the input string. Reanalysis manifests itself in increased processing times (e.g., longer gaze durations, more regressive eye movements).

In constraint-based models, syntactic and conceptual processors analyze the input string *in parallel*, i.e. during one and the same stage. In case of ambiguity, multiple analyses are computed initially and a *competition* is set up between them: The processors evaluate, for each candidate analysis, the extent to which it satisfies relevant constraints of any type—pragmatic, discourse, semantic, morphological, syntactic, phonological, etc. As soon as one candidate reaches an evaluation score that exceeds a threshold, it is deemed the winner and functions as the current (or final) analysis. Typically, the distance between evaluation scores determines the time taken by the competition to subside. This scenario predicts that, on average, competitions between (nearly) well-matched candidates will take longer to settle than competitions between candidates with widely diverging qualities (cf. the duration of tennis or sumo matches). Van Gompel et al. (2005, p. 300) describe the predictions from constraint-based competition models as follows:

> These models predict that when two analyses receive approximately equal support from the various constraints, severe competition should occur, resulting in processing difficulty. In contrast, when one analysis receives much more support than its alternative, there should be little competition and processing should be easy. Crucially, because all sources of information are assumed to have an immediate effect on the activation of the analyses, all sources of information, including semantic plausibility and a word's temporal features, should influence the extent to which the analyses compete. Thus, constraint-based competition models predict that globally ambiguous sentences should be harder to process than [locally] disambiguated sentences.

Consider examples (1) through (3), all from Van Gompel et al. (2005, Experiment 2). Sentence (1) is Globally Ambiguous (GA) because the reduced relative clause *retiring after the troubles* can be plausibly attached both high (to the head noun of the subject NP, i.e. *bodyguard*) and low (to the head noun of the complement of the PP, i.e. *governor*). Variants (2) and (3) are only Locally Ambiguous (LA) as *retiring* fits semantically with only one attachment site—high for (2), low for (3).

(1) I read that the bodyguard of the governor retiring after the troubles is very rich (GA)
(2) I read that the governor of the province retiring after the troubles is very rich (LA-H)
(3) I read that the province of the governor retiring after the troubles is very rich (LA-L)

Van Gompel et al. (2005) collected eye tracking data while the participants were reading sentences like (1), (2) or (3). The key results are presented in Table 1. The relative clause

Table 1 Average reading times (ms) for GA versus LA-L and LA-H sentences (from Van Gompel et al. 2005, Experiment 2)

Ambiguity type	Head verb of relative clause	Remainder of relative clause	Remainder of complement clause	Total of preceding columns
GA	542	797	1065	2404
LA-L	578	880	1103	2561
LA-H	601	899	1073	2573

Notes: Each data point represents the average duration of all fixations that occurred within a region during a given trial. The data columns represent reading times for—from left to right—the head verb of the relative clause (*retiring*), the remainder of the relative clause (*after the troubles*), the last part of the complement clause (*is very rich*), and the sum of the averages in the three preceding columns

appears to be *easier* to read in the GA condition than in the LA conditions. The average reading times (summed over the disambiguating and post-disambiguating regions) amount to 1,339 (GA), 1,458 (LA-L) and 1,500 (LA-H) milliseconds. This pattern of data, which was replicated several times in experiments with different sentence materials, is opposite to what constraint-based models are thought to predict (see above).

The Unrestricted Race Model

Van Gompel et al. (2005) explain their data in terms of a special type of reanalysis model called the Unrestricted Race model. In case of ambiguous input, the parser is assumed to explore, in parallel, all possible analyses of the new input word. The parser may be biased in favor of one or more of the alternative analyses due to prior syntactic as well as non-syntactic information, to syntactic complexity, or otherwise. The prior information includes the meaning computed for the input string up to and including the input word that was processed last. These biases affect the time it takes to complete the alternative analyses. The parser adopts the analysis that finishes first, i.e., wins the speed race, and discards all other ones. If subsequent words cannot be assigned a satisfactory or plausible analysis, the parser is forced to backtrack and to reanalyze (part of) the input string. The model is a syntax-first model because the initial analysis of a novel input word does not take the meaning of this word into account but only its part-of-speech (major word-class) feature. (However, as said before, any biases resulting from prior syntactic and non-syntactic information are supposed to make themselves felt here.)

The sentence materials the participants in Van Gompel et al.'s experiments had to read in the GA condition, had been carefully prepared so as to exclude a strong *a priori* preference for a low or high attachment interpretation. This implies that the processor settles on attaching the verb of the relative clause high and low with roughly equal proportions: The two attachments win the race about equally often. As they both yield a plausible meaning, no reanalysis needs to be carried out. In the two LA conditions, however, when the meaning of the verb is accessed and integrated with the meaning of the current input string (e.g., *province–retiring*), the parser's initial attachment site for the verb of the relative clause turns out to be wrong in about half the trials. Hence, reanalysis will occur in roughly 50% of the trials and manifest itself in prolonged reading times due to—among other things—a higher incidence of regressive eye-movements (eye fixations on earlier text regions; *regressions* for short).

Alternative Interpretations

The Unrestricted Race model is not the only possible interpretation. Consider a slightly—but essentially—different version where the syntactic processor is operating *in parallel* with the non-syntactic processors (cf. Jackendoff (2002) and Kuperberg (2007)). Specifically, the syntax-first assumption of the Unrestricted Race model is dropped and the syntactic races for an attachment site of a new ambiguous input word proceed in parallel with the construction of a plausible conceptual interpretation. The conceptual processor can work in a manner similar to the syntactic processor, except that its decisions are founded on conceptual rather than syntactic constraints and preferences. The conceptual constraints derive from the meanings of the content words, and from the meaning implications entailed by inflections and function words (e.g., the genitive morpheme attached to *governor* in *the governor's province* and the function word *of* in *the province of the governor* are both interpretable as expressing a possessive relationship). Thus, the conceptual processor builds its own hierarchical representation of the current input sentence in parallel to the syntactic representation. In simple cases (e.g., *The dog bit the man*), the first conceptual representation that the conceptual processor proposes without having taken syntactic information into account, will dovetail with the first syntactic representation proposed by the syntactic processor. If in more complex cases, such as *The dog was bitten by the man*, the preferred conceptual analysis happens to conflict with the preferred syntactic parse (cf. Ferreira (2003)), a revision will have to be negotiated between the parallel processors.

In terms of the LA examples (2) and (3), the conceptual processor quickly determines that GOVERNOR, but not PROVINCE, is a plausible actor of RETIRE events. Simultaneously with this conceptual analysis, the race for high versus low attachment of *retiring* takes place within the syntactic processor. After both the syntactic and conceptual processors have finished, a comparison stage follows which checks whether the output structures they produced are compatible. In the two LA-conditions, this yields a 50% probability of a mismatch and mismatch-triggered reanalysis. The predictions from this parallelized version of the Unrestricted Race model coincide with those of the original serial version: In the GA condition, the probability of a mismatch is zero.

A second option adds *interactivity* between syntactic and non-syntactic processing but drops the separate comparison stage. For the sake of simplicity, let us restrict ourselves to two processors—one syntactic, one conceptual—which are dealing with syntactically and conceptually correct input sentences. We make the following three assumptions—the first two about how the processors inform one another, the last one about how eye-movements respond to reanalysis:

1. Whenever a decision has been made by one of the linguistic processors, it is immediately communicated to the other one.
2. A processor that is informed about a decision reached by the other one, immediately checks whether it is compatible with (one of) the option(s) allowed by the current internal state, and discards incompatible options.
3. If, after terminating its work on the current input, a processor is left without acceptable analyses, it initiates reanalysis, which may cause regressive eye movements.

What does this entail with respect to the GA and LA sentences (1) through (3)? First, consider the possible sequences of events during processing GA sentences. There are two possibilities—the conceptual processor reaching a decision prior to the syntactic processor (i.e., it settles on a conceptual attachment without being informed by syntactic information), or vice-versa. Suppose that—for whatever reason—the conceptual processor decides to select

GOVERNOR as the actor of RETIRE, and that this decision reaches the syntactic processor prior to the latter having chosen an attachment site for the relative clause. The syntactic processor then easily verifies that this conceptual choice corresponds to a low attachment site, and immediately adopts this decision, which fits the current syntactic tree. Hence, there is no need for reanalysis by the syntactic parser. The same final conclusion follows when the conceptual processor prefers BODYGUARD as the actor of RETIRING. In cases where the syntactic processor manages to take an attachment decision before the conceptual processor makes its choice, the latter will never fail to adopt the former's. In sum, the GA sentences will never give rise to reanalysis.

The LA sentences may give rise to a different course of events. If the conceptual processor reaches its conclusion ahead of the syntactic processor, the latter will never fail to accommodate this conclusion because the corresponding attachment site is always available. Conversely, if syntactic processing terminates prior to conceptual processing, it may have selected an attachment site that the conceptual processor refuses to adopt—e.g., interpreting PROVINCE as the actor of RETIRING. This does give rise to syntactic reanalysis and regressive eye movements.

How does the model proposed in this Section differ from the constraint-based parallel-interactive models targeted by Van Gompel et al. (2005)? Two aspects stand out. First of all, in current constraint-based models, the processors ("information sources") cannot initialize reanalysis *individually*. Instead, the information recruited by these sources is pooled together ("integrated") and eye movements are affected only after an overall decision has been reached. The second aspect concerns the decision-making mechanism itself, which is based on competition between alternative analyses. If the pooled information clearly supports one of the alternative analyses to the detriment of the other one(s), the winner will emerge sooner than when the evidence is mixed. This mechanism predicts a reading time pattern opposite to what Van Gompel et al. found (but see the unexpected computer simulation results reported by Green and Mitchell 2006[1]). However, the model outlined in the present Section shows that parallelism of syntactic and conceptual processing can be realized in a model where two (or more) concurrent processors are engaged in a race to be the first to complete their jobs. This speed race model is faithful to Van Gompel et al.'s (2005) empirical findings while lacking the "monolithic" pooling of syntactic and conceptual information typical of existing constraint-based models. See the Appendix for a proof of concept.

Competition and Parallel-Interactive Processing Combined

Does the speed race model proposed above rule out competition altogether? The answer is negative: There may very well be competition *within* each (or some) of the linguistic processors. For instance, the selection of a suitable attachment site for the verb *retiring* within the syntactic parser could involve competition instead of a speed race. Specifically, instead of constructing two or more independent parse trees for a syntactically ambiguous input string (as happens in the Unrestricted Race Model), in a competition model an incoming word or word group is tentatively and temporarily attached to all nodes of the current tree that are allowed by the grammar. The ensuing competion between alternative attachment sites is then

---

[1] Green and Mitchell (2006) re-implemented a prominent constraint-based model (McRae et al. 1998) and showed that, at least in this model, the crucial prediction (longer reading times with balanced than with unbalanced alternatives) holds only if the new input word is processed in the absence of biases inherited from earlier input. If, at the onset of processing new ambiguous input, the processor is already biased in favor of one of the alternatives, the predicted effect does not necessarily become manifest.

resolved on the basis of factors such as the frequency or the complexity of the resulting syntactic configurations, the recency of the nodes to which the new input is attached, or their activation level. The details of the mechanism subserving competition within the syntactic processor need not concern us here. Specific proposals have been made by, e.g., Stevenson (1993, 1994), Kempen and Vosse (1989) and Vosse and Kempen (2000). Similarly, within the conceptual processor there may be a competition for the selection of the role of actor of the RETIRE concept.

Conclusion

We have shown that the data reported by Van Gompel et al. (2005) do not qualify as "evidence against competition during syntactic ambiguity resolution." At best, their data count as evidence against the monolithic type of competition operating in current constraint-based models. A "modularized," less monolithic version of a parallel-interactive speed race model of ambiguity resolution accounts for the eye tracking facts. Competition can play an important role in such a model if it is positioned *within* the individual syntactic and conceptual processors.

Appendix

A Parallel-Interactive Race Model for Syntactic and Conceptual Processing During Syntactic Ambiguity Resolution

In this Appendix, we present a stochastic model that accounts for the crucial eye movement data obtained by Van Gompel et al. (2005), Experiments 1 and 2, in particular the "first-pass reading times" and the "first-pass regressions" (Tables 2 and 3 in their paper). The data we use in our calculations are listed in Table A1. The sentences used in Experiment 1 are exemplified in (4) through (6). They differ from those in Experiment 2 (see (1) through (3) above) in that they include a non-reduced relative clause with the head verb *retiring* being preceded by *that will be*.

Table A1 Data from Experiments 1 and 2 by Van Gompel et al. (2005) used to fit the speed race model

	Head verb of relative clause	Remainder of relative clause	Remainder of complement clause
Experiment 1			
First-pass reading time in GA (ms)	312 (8)	574 (14)	861 (24)
First-pass regressions (%)			
in GA	17.7 (2.1)	8.5 (1.5)	58.3 (2.8)
in LA-H	19.2 (2.2)	15.3 (1.9)	61.3 (2.8)
in LA-L	20.9 (2.3)	16.3 (2.0)	64.2 (2.7)
Experiment 2			
First-pass reading time in GA (ms)	378 (10)	552 (16)	851 (22)
First-pass regressions (%)			
in GA	12.1 (2.3)	13.6 (2.3)	63.4 (3.4)
in LA-H	9.5 (2.1)	16.0 (2.5)	64.4 (3.4)
in LA-L	8.4 (2.0)	23.6 (2.9)	69.1 (3.2)

Notes: Standard errors are in parentheses

(4) The bodyguard of the governor that will be retiring after the troubles is very rich (GA)

(5) The governor of the province that will be retiring after the troubles is very rich (LA-H)

(6) The province of the governor that will be retiring after the troubles is very rich (LA-L).

We make the following assumptions:

1. Syntactic processing (S) and Conceptual processing (C) are independent, asynchronous processes with lognormal distributions of their termination times:

$$S = LN(\mu_S, \sigma_S, \theta) \quad \text{and} \quad C = LN(\mu_C, \sigma_C, \theta).$$

For the sake of simplicity, we assume that in both experiments the values of θ (the "offset" or minimal termination time) for S and C are identical. Furthermore, because reduced and non-reduced relative clauses do not differ with respect to the conceptual decisions needed to select the actor for the concept denoted by their head verb (e.g., RETIRE), we assume that the values of μ_C and σ_C in Experiment 1 are identical to those in Experiment 2.

2. Given the sentence materials used in the two experiments, when C terminates before S and relays its decision to S, S will always be able to accept it: zero probability of reanalysis. However, if S terminates before C, it may have selected an analysis that C cannot adopt: a reanalysis probability of 1/2 in the LA conditions, and a zero probability in the GA condition.

3. When C cannot accept the analysis proposed by S (i.e., reanalysis is initiated), a regression ensues with probability p_r. This probability is independent of the state of the system.

4. Regression effects can spill over into subsequent regions (cf. Engbert et al. 2002, for a stochastic model of saccade generation during reading). We combine regressions in all three regions from the head of the relative clause (e.g., *retiring*) until end of sentence. (See Table 1 in the main text for the boundaries between the regions.)

5. The need for reanalysis causes extra regressions over and above the regressions caused by other aspects of the reading process. The occurrence of regressions of the latter type ("normal" regressions) is statistically independent from the occurrence of "reanalysis-driven regressions."

We derive the distributions of S for Experiments 1 and 2 from the first-pass reading times—means of 312 and 378 ms, respectively, and standard deviations of 152 and 179 ms. We decided to work with these numbers despite the fact that first-pass reading times may overestimate the actual syntactic processing times, which also depend on other components of the reading process. Furthermore, the standard deviations may be subject to rounding errors since they were computed from the standard errors of the mean first-pass reading times.

The probability of reanalysis-driven regressions is derived from the probability of regressions (over three regions) in the LA conditions and the corresponding probability in the GA condition: If, in the LA conditions, S terminates earlier than C, reanalysis-driven regressions may occur on top of the "normal" regressions; in the GA condition, all regressions count as "normal." Hence, P(regression in condition LA-X condition) = P(a reanalysis-driven regression in LA-X condition ∨ a "normal" regression in GA condition). Consequently,

$$
\begin{aligned}
&P\,(\text{reanalysis-driven regression in LA-X}) \\
&= \frac{P\,(\text{regression in LA-X}) - P\,(\text{regression in GA})}{1 - P\,(\text{regression in GA})}
\end{aligned}
$$

The probability of a regression in multiple regions is computed by $1 - \prod (1 - p_i)$. For Experiment 1, this gives a 15.7% probability of reanalysis-driven regressions in the LA-H condition and 24.5% for the LA-L condition (an average of 20.1%). For Experiment 2, these figures are 2.6% and 22.2% (12.4% on average).

Reanalysis-driven regressions can only occur when S terminates prior to C. Hence, the probability of one or more extra regressions is equal to half the probability that S terminates before C, times the probability of a reanalysis-driven regression:

$$P \text{ (regression)} = \frac{1}{2} p_r \int_0^\infty P\left(S = t \wedge C \geq t\right) dt$$

The outcome depends on the value of p_r. With $p_r = 1$, a perfect fit (i.e., an exact prediction of the regression probability) results for $\mu_C = 260$, $\sigma_C = 80$ ms, and $\theta = 123$ ms; with $p_r = 0$, no fit is possible.

Figure A1 shows the distributions of S and C and P(reanalysis) for Experiments 1 and 2 on the assumption of a high p_r value (between .8 and 1). The horizontal axis represents the time t needed to process the critical word (e.g., *retiring*). The continuous curve shows the probability density for C terminating at time t; the dotted curve shows the corresponding probability density for S. Although C terminates before S *on average*, it does not always
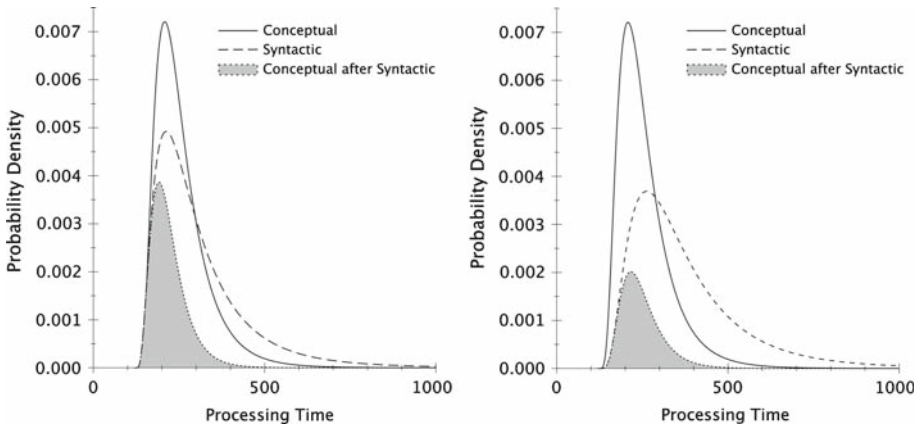

Fig. A1 Distributions of S, C and P(reanalysis) for Experiments 1 (left) and 2 (right)
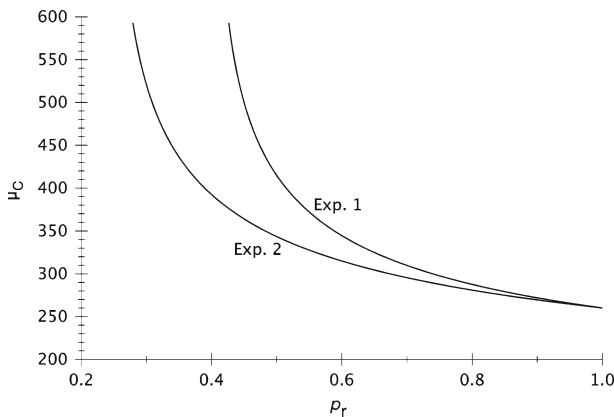

Fig. A2 Values for μ_C that yield a perfect fit of the regression probabilities in Experiments 1 and 2 by Van Gompel et al. (2005), in function of p_r

do so. The "gray" curve shows the probability density of S terminating at t while C has not yet terminated at t. Reanalysis is triggered in (roughly) half of these cases, causing eye movement regression (provided p_r is high). The gray surface represents the probability that S terminates before C (40.2% in Experiment 1 and 24.8% in Experiment 2).

Figure A2 shows which pairs of μ_C and p_r values yield a perfect fit of the regression probabilities in the two experiments. Clearly, the smaller the probability of reanalysis being followed by a regression, the longer C should last in order to enable a perfect fit. In both experiments, when p_r drops below roughly .6, μ_C needs to exceed μ_S (more precisely: below .55 in Experiment 1, and below .61 in Experiment 2). The number of reanalysis-based regressions in Experiments 1 differs considerably from that in Experiment 2, as do the first-pass reading times. Under the assumption that C is about equal in both experiments, our model correctly predicts that the number of regressions increases when the parsing time decreases. In Experiment 1, the context preceding the head verb of the relative clause (e.g., *the bodyguard of the governor that will be*) is syntactically more constraining than in Experiment 2 (*the bodyguard of the governor*). Consequently, the parser can terminate faster in the former than in the latter experiment, thereby also increasing the probability of S terminating prior to C, with a higher probability of reanalysis and regressive eye movements as a result.

References

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42*, 621–636.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*, 164–203.

Green, M. J., & Mitchell, D. C. (2006). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 55*, 1–17.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

Kempen, G., & Vosse, Th. (1989). Incremental syntactic tree formation in human sentence processing: A cognitive architecture based on activation decay and simulated annealing. *Connection Science, 1*, 273–290.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23–49.

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.

Stevenson, S. (1993). A competition-based explanation of syntactic attachment preferences and garden path phenomena. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics (Columbus, Ohio, June 1993)*, pp. 226–273.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research, 23*, 295–322.

Traxler, M. J., Pickering, M. J., & Clifton, C., Jr. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*, 558–592.

Van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 52*, 284–307.

Van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 621–648). Oxford: Elsevier.

Van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language, 45*, 225–258.

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.