UC Davis
UC Davis Previously Published Works

Title

Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing

Permalink

Journal

Trends in Cognitive Sciences, 18(11)

ISSN

1364-6613

Author

Traxler, Matthew J

Publication Date

2014-11-01

DOI

10.1016/j.tics.2014.08.001

Peer reviewed

Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing

Matthew J. Traxler

Department of Psychology, University of California, Davis, 1 Shields Avenue, Davis, CA 95616, USA

Abstract

Syntactic parsing processes establish dependencies between words in a sentence. These dependencies affect how comprehenders assign meaning to sentence constituents. Classical approaches to parsing describe it entirely as a bottom-up signal analysis. More recent approaches assign the comprehender a more active role, allowing the comprehender's individual experience, knowledge, and beliefs to influence his or her interpretation. This review describes developments in three related aspects of sentence processing research: anticipatory processing, Bayesian/noisy-channel approaches to sentence processing, and the 'good-enough' parsing hypothesis.

Keywords

syntax; parsing; Bayes' theorem; noisy channels; good-enough parsing

Syntactic parsing: then and now

Syntactic parsing comprises a set of mental processes that bridges the gap between word-level and discourse-level semantic processes. These interface processes serve to build or recover dependencies between words in a string [1–5] (see [6,7] for the role of syntax and grammar in production). Structural dependencies, conceptual information supplied by content words, and principles governing how thematic role assignments are derived from grammatical functions determine the standard, literal interpretation assigned to a sentence. Take, for instance, the content words embarrass, nurse, and doctor. These words are not sufficient, by themselves, to allow a comprehender to say who did what to whom (or how, when, and where). Syntactic cues and syntactic parsing processes supply the information needed to determine who did what to whom. While the meaning assigned to a given utterance depends on multiple factors, no theory of language understanding can be complete without a consideration of syntax (or grammar) and syntactic parsing.

Psycholinguists have long debated the degree to which syntax and syntactic parsing represent an autonomous, modular subsystem within the larger suite of language production and comprehension processes [8–11]. The prevailing view is that, while there are aspects of syntactic parsing that cannot be subsumed by other levels of processing (e.g., lexical or

Corresponding author: Traxler, M.J. (mjtraxler@ucdavis.edu, mjtraxler@gmail.com).

discourse processes), there are strong interactions between syntactic processes and other aspects of linguistic interpretation (e.g., prior linguistic context, concurrent prosody) and between syntactic processes and aspects of cognition beyond strictly linguistic systems [3,12–15]. Examples of aspects of interpretation that require syntactic computations include phrase packaging (inclusion or exclusion of words from phrases), modifier attachment decisions (in cases where a modifying expression may belong to only one among a set of previous words), and the definition of theta domains and assignment of thematic roles (see Glossary), among others [16–19].

There is general agreement that syntax plays an important role in meaning derivation, but there has been a shift away from strictly bottom-up, serial, encapsulated views of language interpretation and toward more interactive accounts. Three sets of related developments are changing the way that psycholinguists view language interpretation in general and the nature of syntactic parsing processes in particular. These include the relatively recent emphasis on predictive or anticipatory processes, the application of Bayesian probability estimation to language comprehension, and the changing view of language comprehension through the lens of satisficing or good-enough processing [20–25].

The role of anticipation

Language interpretation occurs in a rapid and incremental fashion [26,27]. Comprehenders can identify a word's semantic and syntactic characteristics and the word's relationship to prior context within a few hundred milliseconds of encountering it. Any account that assumes that processing occurs in a strictly bottom-up fashion (signal analysis, followed by word recognition and lexical access, followed by syntactic parsing, followed by integration of new information with prior syntactic and semantic context) is strongly constrained by the speed at which comprehenders can access detailed information about newly encountered words. One way to account for the incredibly rapid and incremental nature of interpretation is to propose that comprehenders anticipate upcoming input rather than waiting passively for the signal to unfold and then reacting to it. Results from various experimental paradigms indicate that comprehenders discriminate between more likely and less likely continuations. In reading, more predictable words are skipped more often than less predictable words [28–30]. Visual world experiments also indicate that comprehenders actively anticipate or predict the imminent arrival of not-yet-encountered information [20]. In these visual world experiments, participants view an array containing pictures of various objects (e.g., a cake, a girl, a tricycle, and a mouse). While viewing the array, participants listen to sentences. If the sentence begins 'The little girl will ride...', participants make eye movements toward the picture of the tricycle even before the offset of the verb 'ride'. This is not simply a reflex based on association between ride and tricycle, however. If the visual array includes a little girl, a man, a tricycle, and a motorcycle, participants make anticipatory eye movements toward the tricycle when listening to 'The little girl will ride...'. However, they make anticipatory movements toward the motorcycle in the same visual array if the subject noun is 'man' (as in 'The man will ride...'). Event-related potential experiments show that prediction-supporting contexts produce smaller N400 responses than less supportive contexts, even when intralexical association is held constant [31].

Results like this indicate that, at least in processing environments where a small number of referents are made visually salient, participants are capable of identifying how a sentence is likely to continue. These predictions may relate to what concept is likely to be mentioned next, but they may be even more specific than that. For example, DeLong and colleagues' event-related potential study [32] produced evidence that comprehenders anticipated the phonological form of an upcoming pair of words (a determiner and a noun). Further, participants appear to act on these predictions before they receive definitive bottom-up evidence confirming or disproving the prediction (i.e., when context makes one continuation more likely than others, the eyes will fixate a picture representing the likely continuation before the comprehender hears a word that refers to that object).

Although researchers agree that anticipation and prediction occur during sentence interpretation, the precise means by which comprehenders derive predictions is currently not well understood. Hence, we need accounts that can tell us how predictions are made, which in turn will tell us why some predictions are made but not others. Hypotheses about how predictions are made include using the production system to emulate the speaker [33,76], relying on intralexical spreading activation [34], or using schematic knowledge of events [35]. A simple word–word association hypothesis is made less plausible by experiments showing that syntactic factors affect the response to a word when lexical association to preceding context is held constant [36]. In addition, words that do not fit a syntactically governed thematic role do not enjoy a processing advantage simply because they are associated with other content words in a sentence [37]. For example, the word axe is strongly associated with the noun lumberjack. Despite this strong association, axe is not processed faster than normal in the sentence frame 'The lumberjack chopped the axe'. Interestingly, the neurophysiological response to a word that is associated with other content words in a sentence changes based on preceding discourse context [31,38–41]. When discourse context activates an event schema that incorporates a particular concept, a word relating to that concept will evoke a smaller N400 response, even when that word is not a good fit given the immediate syntactic context. Hence, event knowledge representations rather than simple lexical co-occurrence appear to provide comprehenders with the basis for deriving predictions.

Bayesian estimation and noisy channels

Comprehenders can anticipate the imminent arrival of specific lexical items relating to activated event representations. Research on syntactic processes suggests that comprehenders may also be able to anticipate structural properties of sentences before bottom-up cues provide definitive evidence for or against a given structural hypothesis. The idea that syntactic parsing processes can be affected by the probability or likelihood of particular syntactic structures has, in fact, been around for a long time [2,3,42]. Trueswell and colleagues were among the first to provide evidence that the conditional probability of a structural analysis in a given context influenced the processing load imposed on the comprehender [10,43]. The precise timing and nature of probabilistic influences are treated differently under different accounts of parsing [19,44–46].

More recent experiments indicate that word recognition is affected by the fit between a word's category distribution profile and the preceding syntactic context [47]. A word that is ambiguous between being a noun or a verb (e.g., card, spin) could be used more often as a noun or as a verb or might appear equally often in each category. Noun-biased words took longer to recognize in a minimal context that favored the verb meaning ('to card'), while the opposite was true of verb-biased ambiguous words ('a spin'). Several studies indicate that lexical statistics attaching to individual words in a sentence (e.g., the frequency with which a verb appears with a direct object versus with a sentence complement) influence the time it takes to parse a sentence. There is also evidence that the syntactic context in which a word appears influences lexical processing (word recognition/lexical decision). The implication is that lexical and syntactic levels of processing place constraints on one another, not that one level takes complete precedence over the other.

The preceding studies indicate that comprehenders anticipate how expressions will continue (in terms of concepts that are likely to be mentioned, the forms that will be used to denote those concepts, and the syntactic structures that are likely to be expressed). Making such predictions entails privileging concepts that are relevant to the described situation and may also entail anticipating the imminent arrival of specific words. Similarly, comprehenders may develop hypotheses that give some syntactic structures advantage over others in an anticipatory fashion. Information that is in some way more predictable or more relevant, whether lexical, conceptual, or structural, must be activated more strongly than other kinds of information. In that case, the comprehender must have some means of determining what is more likely and what is less likely. Recent approaches to sentence processing have incorporated three claims relating to prediction. First, comprehenders do not have a veridical internal representation of the input [48,49]. Second, comprehenders use all available information to compute the likelihood of different interpretations given the cues available in the input [21,24,50–54]. Third, failed predictions are an important factor in the on-line response to input and in the way that knowledge about language is obtained [50]. Surprisal accounts suggest that the information value of a given word in a sentence is a function of its likelihood in context. Words that are relatively less likely will carry more information than words that are more likely. All other things being equal, less likely (more surprising) words lead to greater changes in the knowledge base that is used to derive predictions (see [55,56] for a related approach).

In many cases, the interpretation derived from a 'lowfidelity' representation of the input will match the speaker's intended meaning, but in other cases comprehenders' prior knowledge will lead to systematic distortions in interpretation. These distortions may occur because interpretation does not depend on the signal alone. Interpretation also depends on the comprehender's knowledge about what is likely and what is not likely before the signal arrives. A Bayesian mechanism has to take base-rate information into account when deriving probability estimates. Comprehenders integrate base-rate information (how likely is a given interpretation in the absence of any evidence) with information available in the stimulus to rank interpretations from more likely to less likely. If the signal is noisy or is conveyed over a noisy channel, interpretation will be systematically biased toward higher-frequency interpretations.

Gibson and colleagues' formulation of this type of account introduces the idea that the degree of perceived noise, and assumptions about the kinds of distortion that are likely to occur because of noise, affect the interpretations that comprehenders assign to sentences [24]. Given that the transmission channel is noisy, a comprehender may assume that the signal is missing a component. Noise need not be purely environmental. 'Internal noise' can be caused by inattention, distraction, fatigue, or boredom. These internal factors influence the rate of uptake of information and the nature and quality of long-term memory for language input [57–59].

Comprehenders are less likely to assume that something has been spuriously added to the signal. This leads to predictions about the way that comprehenders will respond to specific types of sentence that (if the signal is perceived accurately) are anomalous. For example, if a person heard or read 'The mother gave the candle the girl', this is less disruptive than hearing or reading 'The mother gave the girl to the candle' (although the licensed interpretation is the same in either case). In the former case, the sentence can be 'repaired' by assuming that noise wiped out the part of the signal between 'candle' and 'the'. Noise is a less plausible explanation for the anomaly in the second sentence. Gibson and colleagues showed that a metric of deviance based on the number and types of edits required to make a semantically anomalous sentence plausible provided accurate estimates of the extent to which participants would rate an anomalous sentence as plausible. These differences in acceptability are not predicted by an account under which comprehenders always perceive the sentence accurately.

Similar assumptions may account for other syntactic complexity effects [60] and similarity-based interference effects [60,61]. Memory-based interference relies on the idea that mental representations can interfere with one another and that more similar mental representations interfere with one another more than less similar representations. This resembles the systematic misperception that is assumed to underlie interpretive errors by noisy-channel accounts. 'Doctor' and 'lawyer' may be difficult to distinguish at a semantic level in the same way that 'at' and 'as' are difficult to differentiate at the phonological level. The latter problem has been claimed by noisy channel proponents to account for processing phenomena in a certain type of difficult, 'garden-path' sentence [62]. Structures that impose greater memory demands may be more susceptible to disruption by noise (internal or external). Similarly, noise may have greater effects when two representations are more similar in terms of their phonology (e.g., nurse–purse) or meaning (e.g., doctor–lawyer) than when they are more distinct (nurse– handbag, doctor–hamster) [63].

Challenges for noisy-channel models

While Bayesian estimation and noisy-channel hypotheses account for various empirical phenomena, it remains unclear whether accounts derived for specific sentence types are computationally tractable for natural language more broadly. This is due to the combinatorial explosion that occurs for any system that must compute interpretations, both for the signal and for near neighbors across different dimensions (phonological, lexical, syntactic) and different grain sizes [64]. Bayesian approaches to parsing assume that a comprehender must compute and evaluate not just the syntactic structures and interpretations

that can be assigned to an accurately perceived string, but also all of the near neighbors that would be produced by edits to the string. Given the number of lexical, structural, and semantic replacements that are possible, the number of computed interpretations would be astronomical. However, there may be ways of limiting the search space such that the space consistently includes the intended interpretation without requiring the computation and evaluation of massive numbers of alternative interpretations [65].

Levy [49] noted that Bayesian noisy-channel models faced the challenge of predicting and explaining a wide variety of experimental results across a wide variety of tasks and sentence types. Bayesian noisy-channel models may resist being scaled up for a few reasons. Currently, much of the empirical evidence for noisy-channel models comes from sentence judgments, a kind of 'off-line' task. Off-line tasks are those that measure the consequences of comprehension rather than the process of comprehension itself. A full-coverage account should be able to handle the results of both off-line and on-line performance. Second, it is unclear why a system that is indifferent to the precise nature of the signal would reliably and very rapidly detect ungrammaticalities. However, as Phillips and Lewis note, comprehenders 'detect just about any linguistic anomaly within a few hundred milliseconds of the anomaly appearing in the input' ([66], p. 19). Third, some of the on-line evidence for the noisy-channel hypothesis comes from very restricted processing environments. For example, Levy and colleagues' paper on reduced relative clause processing focused on the possibility that comprehenders might misperceive the word 'at' as the word 'as' [62]. Comprehenders frequently misinterpret sentences such as 'The coach smiled at the player tossed the frisbee' as meaning that the player tossed the frisbee (the grammar licenses the meaning that somebody tossed the player a frisbee). The noisy-channel model attributes this misinterpretation to misperception of the sentence (people hear it as 'The coach smiled *as* the player tossed the frisbee'). In the 2009 experiment, all of the experimental sentences had reduced relative clauses in object position and they all were preceded by the word 'at'. However, processing difficulty and misinterpretation could remain even if the word 'at' were replaced by various other words (smiled beside..., smiled in front of..., smiled longingly toward...) that are not as confusable with 'as', although the relevant experiment has not been conducted. Finally, as Pylkkannen and McElree note, any kind of underspecification account, including noisy-channel models, may be unable to explain how novel meanings are constructed in the first place ([67], p. 540). Interpretive systems whose main function is pattern recognition do not, without some major additions, have the ability to assign meanings to unfamiliar patterns. It is also probably not viable to assume syntactic underspecification in production, since speakers must ultimately select one set of morphosyntactic and phonological forms for any given sentence [6,7]. Comprehension and production need not operate in exactly the same way, but there is substantial evidence for shared representation and similar if not identical syntactic structure-building processes across production and comprehension [68].

Satisficing

Bayesian/noisy-channel approaches to parsing hypothesize that interpretation may be based not on the produced signal but on a comprehender's approximation of the signal. This approximation depends both on cues extracted from the signal and on prior beliefs about

events, what speakers are likely to say about those events, and how they are likely to structure their expressions. These assumptions are largely if not fully compatible with a strand of research based on the good-enough parsing hypothesis [22,23,69]. Ferreira and colleagues discovered that comprehenders often derived meanings that were incompatible with the grammatically licensed interpretations of temporarily syntactically ambiguous sentences as well as sentences that were unambiguous. For instance, participants would interpret 'While Mary was dressing the baby played on the floor' as meaning that Mary was dressing the baby. This meaning is not licensed by the signal (the licensed meaning is that Mary was dressing herself). Comprehenders also systematically misinterpret semantically anomalous sentences (sentences for which the licensed meaning is implausible), especially when the syntactic structure is less frequent or noncanonical. Comprehenders will tend to interpret the passive-voice sentence 'The mouse was eaten by the cheese' as meaning that the mouse ate the cheese. This tendency is stronger than for active-voice sentences such as 'The cheese ate the mouse' (Figure 1). There is also some evidence that comprehenders do not resolve syntactic ambiguities involving modifier attachments unless specifically prompted to do so by comprehension questions [70], although this effect differs cross-linguistically and may depend on comprehenders' working-memory capacity [71,72]. Because comprehenders' interpretations in these studies were systematically biased toward plausible meanings, these findings offer a clear demonstration of the effects of prior knowledge and beliefs. That is, they indicate that anomalous interpretations are not adopted even when the syntax of a sentence provides clear indications that the anomalous meaning was intended. Thus, the results are straightforwardly compatible with the Bayesian/noisy-channel approach to sentence interpretation.

These results do not, by themselves, discriminate between two hypotheses regarding syntactic and semantic analysis. The first is that comprehenders computed syntactic form as dictated by the available cues, derived semantic interpretations from syntactic form, and then rejected those meanings because they conflicted with prior beliefs (i.e., mice eat cheese, not the other way around). The second is that semantic interpretations were derived from the lexical content of the sentences without the benefit of syntactically driven thematic role assignment. This latter syntactic underspecification hypothesis would be compatible with the noisy-channel proposal (i.e., comprehenders derive an approximate set of structures for a given string). However, recent work involving comprehenders' response to reflexive pronouns calls the syntactic underspecification hypothesis into question [73]. While reading sentences similar to 'After the banker phoned, Steve's father/mother grew worried and gave himself five days to reply', readers showed a gender-mismatch effect on the reflexive pronoun 'himself'. That means that readers spent more time when the pronoun (himself) did not match the gender of the preceding noun (mother).

This mismatch effect should occur only if the comprehenders have built a detailed syntactic structure for the sentence (based on previous studies on the effects of sentence structure on co-reference) [14,61]. These results suggest that comprehenders do not underspecify syntactic form, but that plausible semantic interpretations derived early in a sentence are not always displaced based on later processing events [74–76]. Hence, while comprehenders may favor sensible interpretations over less sensible ones (based on general world knowledge or contextually supplied information), this does not mean that syntactic

structures are not computed or do not contribute to sentence interpretations. If this is correct, the noisy-channel assumption of underspecified syntax may need to be modified. If the assumption of multiple syntactic structure activations, weighted by frequency and similarity to the signal, were discarded, a different explanation for editdistance effects would be required. Perhaps such effects could result from the kinds of recovery process that are presumed by accounts of syntactic reanalysis or self-monitoring and self-correction during speech production. According to this kind of account, comprehenders would build and assess one syntactic structure at a time. When that structure led to a faulty interpretation, comprehenders would undertake syntactic structure revisions or word-based edits (e.g., 'cheese' is edited to 'chief') to bring the word-level, syntactic, and sentence-level interpretations into agreement.

Open questions

The Bayesian/noisy-channel hypothesis answers several questions about how predictions are drawn. For instance, it indicates that the conditional probability of various syntactic interpretations is computed by combining base-rate information with cue-driven estimation of likelihood. Other questions about how predictions are drawn remain unanswered. Are predictions drawn in a graded fashion or does prediction work like implicit lexical selection, with one candidate ultimately selected? Does prediction depend on emulating the speaker or writer? If so, one would expect that making comprehenders undertake covert articulation would disrupt the prediction process. Another set of questions revolves around the extent to which free attentional and working-memory resources are required for prediction. If prediction depends on free resources, one would expect to see individual differences. Comprehenders with better executive control or more working-memory capacity should make more predictions and predict with greater success than comprehenders with fewer resources. Individual differences may also be found in the way that comprehenders acquire the knowledge that drives estimates of prior probability. Given that the Bayesian account appeals to baserate information as a factor that affects prediction, different comprehenders with different exposure to language input should derive different predictions or at least draw the same predictions with different degrees of certainty. Given that event knowledge feeds into predictions [39], comprehenders with different degrees of expertise in a given content area should derive different predictions, based on their different conceptualizations of the domain. Here again, resources may be an important factor. If gathering of priors is cost free and automatic, mere exposure to patterns in the language should lead to prior probability estimates that closely match the languagewide patterns. If, however, adjustment of priors depends on a mismatch between predicted and observed input, if generating the prediction requires speaker emulation, and if all of that requires some degree of attention, mere exposure will not lead to acquisition of base-rate information or changes in existing estimates of prior probability. If so, task-related factors influencing depth of processing may play an important role in the acquisition and modification of base-rate information [70,77]. All of these questions will require careful investigation before claims can be made with confidence.

Concluding remarks

Classical accounts of syntactic parsing present it as a largely reactive process. According to these classical views, words are identified, syntactic dependencies are computed, and structured sequences are subjected to semantic analysis. This bottom-up, encapsulated view of processing does not offer elegant or clear explanations for experimental results such as Altmann and Kamide's findings that attention shifts to concepts that do not yet have a counterpart in the signal [20]. Alternative approaches embrace the notion that comprehenders use prior knowledge to predict how a string is likely to unfold, in terms of both its lexical and semantic content and its syntactic structure. The noisychannel hypothesis elaborates on this approach by focusing on the kinds of assumption that comprehenders have about signal distortion. Bayesian approaches, like prior constraint-based lexicalist approaches to parsing, assume that parsing a string entails simultaneously computing and evaluating a set of structural and semantic alternatives. The good-enough parsing hypothesis resembles Bayesian/noisy-channel accounts in that it supposes that comprehenders' prior beliefs strongly influence the interpretation derived for a given sentence. This supposition is borne out by findings that comprehenders prefer plausible, illicit interpretations over implausible, grammatically licensed interpretations.

Acknowledgments

Glossary

Event-related potentials (ERPs)
when neurons fire, they generate electrical current that can be detected at the scalp. Neural activity produces systematic oscillations of electrical current. This activity occurs in response to various stimuli. Presentation of a stimulus leads to an ERP – the pattern of electrical activity at the scalp that occurs because of the stimulus.

Grain size
probabilistic accounts, including Bayesian/noisy-channel accounts, suppose that people keep track of patterns in the language. These patterns can occur at different levels of specificity. For example, the most common structure in English is noun–verb–noun (NVN) (subject–verb–object). That is a very large grain size. However, a verb like sneeze almost never takes a direct object. It is most often expressed as noun–verb (NV). So, at a fine grain size, sneeze is most likely to appear in a NV structure. At a larger grain size, the most likely structure is NVN.

Lexical co-occurrence
some words are more likely to appear together than others; police–car is more likely than police–cat.

Lexical item/lexical processing

a lexical item is roughly a single word. Lexical processing refers to the mental operations that retrieve or activate stored knowledge about words as needed during comprehension and production.

N400 response

the N400 is a characteristic of brain waves. The amplitude (size) of the N400 is related to how frequently a word appears in the language, how well the word's meaning fits with its contexts, and other factors that make identifying and integrating the word easier or more difficult.

Reduced relative clauses

relative clauses are expressions that modify preceding nouns. They are often signaled by a relativizer, a word like 'that'. In the expression 'The cat that my sister likes', 'that my sister likes' is a relative clause modifying cat. If the word 'that' were removed, the relative clause would be a reduced relative clause.

Schema

a knowledge structure in long-term memory that reflects an individual's knowledge of a certain kind of event. For example, a restaurant schema encodes the typical participants, objects, and events that occur when one goes to a restaurant.

Sentence complement

sometimes verbs appear with an entire sentence as an argument (as opposed to, for example, just a noun phrase). In the sentence 'John knows the answer is in the book', 'the answer is in the book' is a sentence complement that is governed by the verb 'knows'. (What does John know? That the answer is in the book.)

Syntactic parsing

involves the set of mental operations that detects and uses cues in sentences to determine how words relate to one another.

Syntactic structure

a mental representation that captures dependencies between words in sentences.

Thematic roles

abstract semantic classes that capture common roles that many different entities and objects play in different sentences. For example, in the sentences 'John drank milk' and 'John pokes bears', John is the initiator of the described action. John is a thematic agent. In the two sentences, milk and bears are on the receiving end of the action, so they are thematic patients.

Theta domains

according to Frazier and Clifton [16] some words in sentences assign thematic roles to other words in the sentence. A theta domain is that part of a sentence for which a given word assigns thematic roles.

References

1. Frazier L (1979) On Comprehending Sentences: Syntactic Parsing Strategies, Indiana University Linguistics Club

2. Ford M et al. (1982) A competence based theory of syntactic closure In The Mental Representation of Grammatical Relations (Bresnan JW, ed.), pp. 727–796, MIT Press

3. MacDonald MC et al. (1994) Lexical nature of syntactic ambiguity resolution. Psychol. Rev 101, 676–703 [PubMed: 7984711]

4. Pickering MJ and van Gompel RPG (2006) Syntactic parsing In The Handbook of Psycholinguistics (2nd edn) (Traxler MJ and Gernsbacher MA, eds), pp. 455–503, Elsevier

5. Traxler MJ (2012) Introduction to Psycholinguistics: Understanding Language Science, Wiley–Blackwell

6. Ferreira V and Griffin Z (2006) Properties of spoken language production In The Handbook of Psycholinguistics (2nd edn) (Traxler MJ and Gernsbacher MA, eds), pp. 21–60, Elsevier

7. Levelt WJM (1989) Speaking: From Intention to Articulation, MIT Press

8. Altmann GTM and Steedman MJ (1988) Interaction with context during human sentence processing. Cognition 30, 191–238 [PubMed: 3215002]

9. Fodor JA (1983) Modularity of Mind, MIT Press

10. Trueswell JC et al. (1993) Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. J. Exp. Psychol. Learn. Mem. Cogn 19, 528–553 [PubMed: 8501429]

11. Clifton C et al. (2003) The use of thematic role information in parsing: syntactic processing autonomy revisited. J. Mem. Lang 49, 317–334

12. Fedorenko E et al. (2009) Structural integration in language and music: evidence for a shared system. Mem. Cogn 37, 1–9

13. Schafer AJ et al. (2005) Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task In Approaches to Studying World-situated Language Use: Psycholinguistic, Linguistic and Computational Perspectives on Bridging the Product and Action Tradition (Tanenhaus M and Trueswell J, eds), pp. 209–225, MIT Press

14. Scheepers C et al. (2011) Structural priming across cognitive domains: from simple arithmetic to relative-clause attachment. Psychol. Sci 22, 1319–1326 [PubMed: 21873569]

15. Tanenhaus MK et al. (1995) Integration of visual and linguistic information in spoken language comprehension. Science 268, 1632–1634 [PubMed: 7777863]

16. Frazier L and Clifton C Jr (1996) Construal, MIT Press

17. Frazier L and Fodor JD (1978) The sausage machine: a new two-stage parsing model. Cognition 6, 291–325

18. Jackendoff R (2002) Foundations of Language, Oxford University Press

19. Traxler MJ et al. (1998) Adjunct attachment is not a form of lexical ambiguity resolution. J. Mem. Lang 39, 558–592

20. Altmann GTM and Kamide Y (1999) Incremental interpretation at verbs: restricting the domain of subsequent reference. Cognition 73, 247–264 [PubMed: 10585516]

21. Brown M et al. (2012) Syntax encodes information structure: evidence from on-line reading comprehension. J. Mem. Lang 66, 194–209

22. Ferreira F et al. (2002) Good-enough representations in language comprehension. Curr. Dir. Psychol. Sci 11, 11–15

23. Ferreira F and Patson ND (2007) The 'good enough' approach to language comprehension. Lang. Linguist. Compass 1, 71–83

24. Gibson E et al. (2013) Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. Proc. Natl. Acad. Sci. U.S.A 110, 8051–8056 [PubMed: 23637344]

25. Mahowald K et al. (2012) Info/information theory: speakers choose shorter words in predictive contexts. Cognition 126, 313–318 [PubMed: 23116925]

26. Marslen-Wilson WD (1973) Linguistic structure and speech shadowing at very short latencies. Nature 244, 522–523 [PubMed: 4621131]

27. Marslen-Wilson WD and Tyler LK (1975) Processing structure of sentence perception. Nature 257, 784–786 [PubMed: 1186856]

28. Rayner K (1998) Eye movements in reading and information processing: 20 years of research. Psychol. Bull 124, 372–422 [PubMed: 9849112]

29. Rayner K and Pollatsek A (1989) The Psychology of Reading, Erlbaum

30. Choi W and Gordon PC (2014) Word skipping during sentence reading: effects of lexicality on parafoveal processing. Atten. Percept. Psychophys 76, 201–213 [PubMed: 24170376]

31. Lau EF et al. (2013) Dissociating N400 effects of prediction from association in single-word contexts. J. Cogn. Neurosci 25, 484–502 [PubMed: 23163410]

32. DeLong K et al. (2005) Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nat. Neurosci 8, 1117–1121 [PubMed: 16007080]

33. Pickering MJ and Garrod S (2013) An integrated theory of language production and comprehension. Behav. Brain Sci 36, 329–347 [PubMed: 23789620]

34. Schwanenflugel PJ and LaCount KL (1988) Semantic relatedness and the scope of facilitation for upcoming words in sentences. J. Exp. Psychol. Learn. Mem. Cogn 14, 344–354

35. Schank R and Abelson RP (1977) Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures, Erlbaum

36. Morris RK (1994) Lexical and message level sentence context effects on fixation times in reading. J. Exp. Psychol. Learn. Mem. Cogn 20, 92–103 [PubMed: 8138791]

37. Traxler MJ et al. (2000) Processing effects in sentence comprehension: effects of association and integration. J. Psycholinguist. Res 29, 581–595 [PubMed: 11196064]

38. Camblin CC et al. (2007) The interplay of discourse congruence and lexical association during sentence processing: evidence from ERPs and eye tracking. J. Mem. Lang 56, 103–128 [PubMed: 17218992]

39. Metusalem R et al. (2012) Generalized event knowledge activation during online sentence comprehension. J. Mem. Lang 66, 545–567 [PubMed: 22711976]

40. Hess DJ et al. (1995) Effects of global and local context on lexical processing during language comprehension. J. Exp. Psychol. Gen 124, 62–82

41. Traxler MJ and Foss DJ (2000) Effects of sentence constraint on priming in natural language comprehension. J. Exp. Psychol. Learn. Mem. Cogn 26, 1266–1282 [PubMed: 11009257]

42. Elman J (1991) Distributed representations, simple recurrent networks, and grammatical structure. Mach. Learn 17, 195–225

43. Trueswell JC et al. (1994) Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. J. Mem. Lang 33, 285–318

44. Pickering MJ et al. (2000) Ambiguity resolution in sentence processing: evidence against frequency-based accounts. J. Mem. Lang 43, 447–475

45. Pickering MJ and Traxler MJ (2003) Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. Lang. Cogn. Process 18, 469–473

46. Van Gompel RPG et al. (2001) Reanalysis in sentence processing: evidence against current constraint-based and two-stage models. J. Mem. Lang 45, 225–258

47. Fedorenko E et al. (2012) The interaction of syntactic and lexical information sources in language processing: the case of the noun–verb ambiguity. J. Cogn. Sci 13, 249–285

48. Levy R (2008) Expectation-based syntactic comprehension. Cognition 106, 1126–1177 [PubMed: 17662975]

49. Levy R (2008) A noisy-channel model of rational human sentence comprehension under uncertain input. In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, pp. 234–243. Association for Computational Linguistics

50. Hale JT (2001) A probabilistic early parser as a psycholinguistic model. In Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics, pp. 159–166. Association for Computational Linguistics

51. Hale JT (2011) What a rational parser would do. Cogn. Sci 35, 399–443

52. Jaeger FT and Snider NE (2013) Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. Cognition 127, 57–83 [PubMed: 23354056]

53. Levy R et al. (2013) The processing of extraposed structures in English. Cognition 122, 12–36

54. Smith NJ and Levy R (2013) The effect of word predictability on reading time is logarithmic. Cognition 128, 302–319 [PubMed: 23747651]

55. Chang F et al. (2006) Becoming syntactic. Psychol. Rev 113, 234–272 [PubMed: 16637761]

56. Chang F et al. (2012) Language adaptation and learning: getting explicit about implicit learning. Lang. Linguist. Compass 6, 259–278

57. Reichle ED et al. (2010) Eye movements during mindless reading. Psychol. Sci 21, 1300–1310 [PubMed: 20679524]

58. Sayette MA et al. (2010) Out for a smoke: the impact of cigarette craving on zoning out during reading. Psychol. Sci 21, 26–30 [PubMed: 20424018]

59. Schooler JW et al. (2011) Meta-awareness, perceptual decoupling, and the wandering mind. Trends Cogn. Sci 15, 319–326 [PubMed: 21684189]

60. Gibson E (1998) Linguistic complexity: locality of syntactic dependencies. Cognition 68, 1–76 [PubMed: 9775516]

61. Gordon PC et al. (2001) Memory interference during language processing. J. Exp. Psychol. Learn. Mem. Cogn 23, 381–405

62. Levy R et al. (2009) Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. Proc. Natl. Acad. Sci. U.S.A 106, 21086–21090 [PubMed: 19965371]

63. Acheson DJ and MacDonald MC (2011) The rhymes that the reader perused confused the meaning: phonological effects during on-line sentence comprehension. J. Mem. Lang 65, 193–207 [PubMed: 21743771]

64. Mitchell DC (1987) Lexical guidance in human parsing: locus and processing characteristics In Attention and Performance XII (Coltheart M, ed.), pp. 601–618, Erlbaum

65. Dubey A et al. (2013) Probabilistic modeling of discourse-aware sentence processing. Top. Cogn. Sci 5, 425–451 [PubMed: 23616488]

66. Phillips C and Lewis S (2013) Derivational order in syntax: evidence and architectural consequences. Stud. Linguist 6, 11–47

67. Pylkkannen L and McElree BD (2006) The syntax–semantics interface: on-line composition of sentence meaning In The Handbook of Psycholinguistics (2nd edn) (Traxler MJ and Gernsbacher MA, eds), pp. 537–577, Elsevier

68. Tooley KM and Bock JK (2014) On the parity of structural persistence in language production and comprehension. Cognition 132, 101–136 [PubMed: 24803423]

69. Christianson KA et al. (2001) Thematic roles assigned along the garden path linger. Cogn. Psychol 42, 368–407 [PubMed: 11368528]

70. Swets B et al. (2008) Underspecification of syntactic ambiguities: evidence from self-paced reading. Mem. Cogn 36, 201–216

71. Swets B et al. (2007) The role of working memory in syntactic ambiguity resolution: a psychometric approach. J. Exp. Psychol. Gen 136, 64–81 [PubMed: 17324085]

72. Traxler MJ (2010) A hierarchical linear modeling analysis of working memory and implicit prosody in the resolution of adjunct attachment ambiguity. J. Psycholinguist. Res 38, 491–509

73. Slattery TJ et al. (2013) Lingering misinterpretations of garden-path sentences arise from competing syntactic representations. J. Mem. Lang 69, 104–120

74. Christianson K et al. (2010) Effects of plausibility on structural priming. J. Exp. Psychol. Learn. Mem. Cogn 36, 538–544 [PubMed: 20192548]

75. Pickering MJ and Traxler MJ (2001) Strategies for processing unbounded dependencies: lexical information and verb–argument assignment. J. Exp. Psychol. Learn. Mem. Cogn 27, 1401–1410 [PubMed: 11713875]

76. Pickering MJ and Traxler MJ (1998) Plausibility and recovery from garden paths: an eye-tracking study. J. Exp. Psychol. Learn. Mem. Cogn 24, 940–961

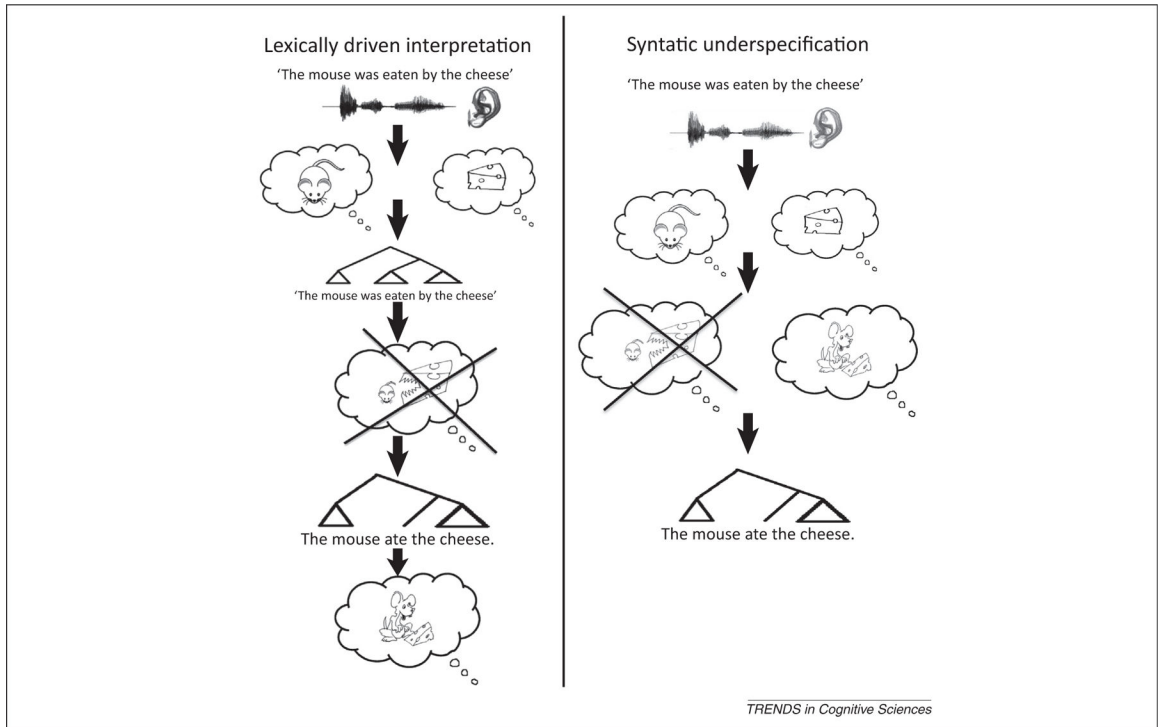77. Christianson K and Luke S (2011) Context strengthens initial misinterpretations of text. Sci. Stud. Read 15, 136–166

Figure 1.
A processing architecture under which default lexical information overrides syntactically derived meanings appears on the left side. Under these assumptions, syntax is computed and then ignored in favor of lexically derived meanings. The right side shows how a similar interpretive outcome could result if syntactic computations were foregone altogether.