Connectedness: Papers by and for Sarah VanWagenen
UCLA Working Papers in Linguistics 18: 259–273
Carson T. Schütze & Linnaea Stockall (eds.), 2014

Memoization in top-down parsing*

Analysis, attention, memory and reanalysis

Edward P. Stabler

There is abundant evidence that ordinary fluent speech is typically analyzed and interpreted on a word-by-word or even syllable-by-syllable basis. This suggests that the structure building in language comprehension is, at least to a good first approximation, top-down and left-to-right (i.e., temporally 'beginning-to-end'). However, standard top-down backtrack parsing is intractable, and practical experiments confirm that it is infeasible for structures that people have no difficulty with. Restricting top-down parses to a beam of probable alternatives would solve the problem if comprehensible utterances were relatively probable so that they stayed in the beam, but, fairly often, they aren't and they don't. The natural alternative is to consider dynamic programming, all-paths, 'memoizing' parsing strategies, but these are disfavored by many linguists and psycholinguists. This paper provides a *prima facie* defense of such models. They are, in a clear sense, top-down, left-to-right and incremental. By remembering ('memoizing') intermediate results, intractability can be avoided, and certain 'reanalysis' effects are easily accommodated. Reanalysis in this model is not the use of special mechanisms for tree repair, but rather a standard use of grammatical rules to construct alternatives from parts already computed.

Keywords parsing, syntax, reanalysis, memoization

Basic and relatively uncontroversial assumptions about the structures of human language together with clear facts about the use of human language pose fundamental problems that we do not yet know how to resolve. Human languages seem capable of exhibiting complexity that is 'properly mildly context sensitive' in the sense of being beyond the (weak and strong) expressive power of context free grammar but within the power of mildly context sensitive grammar (Joshi 1985). And both language comprehension and language production seem to be 'incremental' in surprising respects. That is, the structures are calculated from beginning-to-end in time, with remarkably agility. These two basic assumptions are briefly reviewed because they suggest that an adequate model of human language understanding requires a top-down, left-to-right mildly context sensitive parser. But this raises a puzzle: standard top-down parsers, building connected parses one at a time according to the grammar, are not tractable in general and not feasible as models of what people do. One standard response to that infeasibility has been to build a 'beam' of probable parses (Jurafsky 1996:for example).

*Thanks to Sarah VanWagenen for many stimulating discussions, shaped by her unflagging attention to fundamentals, and to Janet Fodor for insights into some trends of recent theory.

And while that kind of model can be extended for mildly context sensitive grammars (Stabler 2013), it is not a plausible model of human performance. After quickly reviewing the basic assumptions and the problems facing standard top-down backtrack and beam models, a different kind of model is sketched here and defended against some of the criticisms that have kept it from serious consideration. We defend the hypothesis that human parsers are Earley-like, exhaustive, all-paths-at-once, top-down, memoizing parsers.

0 Basic assumptions

While there is an unsurprising diversity of opinion about the structure of language and about our abilities to produce and understand those structures, here we need only a couple of uncontroversial assumptions.

Mild context sensitivity. Joshi (1985) and his collaborators noticed a surprising convergence of a number of independently proposed grammar formalisms in the 1980's. There are actually two points of convergence, supporting two versions of the claim that human languages are mildly context sensitive (MCS). The weaker (empirically safer) claim is that human languages are defined (weakly and strongly) by a range of formalisms that include (set local) multiple component tree adjoining grammars (MCTAGs), a kind of generalized phrase structure grammar called multiple context free grammar (MCFGs), and minimalist grammars (MGs).[2] The main challenges to this weaker proposal argue that even more powerful mechanisms may be needed (Kobele 2006; Michaelis and Kracht 1997; Rambow 1994; Manaster-Ramer 1986), but as Jäger and Rogers (2012) say, "Most experts... assume at this time that all natural languages are MG languages."

Note that this claim does not entail anything about the minimum complexity of particular human languages; it is compatible even with the view that there are finite human languages.[3] But the much stronger claim that human languages are (weakly and strongly) context free has proved untenable, and the failure of that hypothesis provides strong evidence for the view that that the analytic mechanisms available to human speakers are powerful enough to acquire and use languages that are not context free. So if Joshi is right, we have the corollary that context free grammars (CFGs), generalized phrase structure grammars (GPSGs), and filler-gap parsers (formalized, for example, by linear indexed grammars) are inadequate because they are empirically too weak, while the *Aspects* grammar (Peters and Ritchie 1973), augmented transition networks (ATNs), head driven phrase structure grammar (HPSG), and lexical-functional grammar (LFG) are relatively uninteresting and non-explanatory because they are too strong.

Incremental interpretation. It is not quite clear to what extent our perceptions of linguistic stimuli can be shaped by our expectations, by some predictive aspects our linguistic

[2] See Harkema (2001a); Michaelis (2001); Seki et al. (1991); Weir (1988) and references there.

[3] In fact, this is a view I hold. I believe that certain early stages of child languages are best characterized by grammars that define only finite languages. They are nevertheless importantly similar to normal, adult languages and hence deserve to be called human languages. It is natural to say that this is so even for the empty grammar, the grammar that does not yet have the materials to produce any sentence of any language but only the potential of doing so, a potential given by the mechanisms of 'universal grammar'.

abilities,[4] but there is abundant confirmation of the suggestion of Marslen-Wilson (1975) that people typically analyze and understand ordinary fluent speech (or read and understand simple text) within a few hundred milliseconds – a little more than the duration of a typical syllable.[5] This suggests that, at least in the materials of ordinary fluent conversation, a rapid analysis can compute grammatical relations between each word and preceding material, as it is heard.

1 Top-down backtrack and beam models

Among the canonical stack-based, one-parse-at-a-time parsing methods, only top-down parsing is guaranteed to integrate each new word into the parse (the tree of grammatical relations) as soon as it is heard, setting the stage for interpretive mechanisms and inference. Left-corner and bottom-up parsers do not have this property.

It has been difficult to confirm independently that human parsing is top-down. This is one of the points noted by VanWagenen et al. (2011, 2014). Roughly speaking, top-down methods do more work at the beginning of a phrase, while bottom-up methods do more work at the end of a phrase. Finding evidence to bear on which is actually happening is difficult partly because ends and beginnings of phrases are often adjacent. When they are not adjacent – as for example at the end of an utterance, where we have an ending of a phrase with no adjacent beginning – other things happen that obscure the structural complexity. However, top-down models remain most plausible because they construct interpretable grammatical relations to each incoming word.

Standard top-down parsing models are not tenable, though. Far from allowing speedy human-like analysis, they can actually be non-terminating if there is recursion in left branches. And if anything like current syntax is on the right track, human languages all have recursive left branches. Furthermore, even if there were no recursive left branches, top-down backtrack parsing can be intractable in the presence of extensive ambiguity (Aho and Ullman 1972:§4.1), which human languages also have (Church and Patil 1982).

These complexity problems can, in principle, be avoided if the parser can be restricted to a finite 'beam' of desired parses. For example, when parses are associated with locally determined probabilities, as in probabilistic context free grammars (PCFGs) and their extensions to mildly context sensitive (MCS) grammars (Hunter and Dyer 2013), a beam could be restricted to the k most-probable parses, for some particular finite k. An MG parser of this kind is proposed by Stabler (2013). But in these parsers, if the finite bound k is removed, the search for parses is just the standard top-down search and so can fail to terminate when the grammar has left recursion. And on the other hand, if a reasonably small k is imposed, one finds that desired parses fall outside the beam and are missed. For some practical engineering applications, missing a small fraction of the desired parses may not be a problem,[6] but for models of human abilities, this failure indicates a serious flaw in the approach. It is difficult to clinch this point decisively, since one can always hope for a better (but still efficiently definable) probability distribution, or let the beam be a little wider. But

[4] Norris et al. (2000); Tanenhaus et al. (2000).

[5] Staub (2011); Tanenhaus and Brown-Schmidt (2008); Clifton et al. (2006); Sturt (2004); Chambers et al. (2002); Altmann and Kamide (1999); Frazier (1999).

[6] Cf., e.g. Roark (2001, 2004).

people seem to be able to recognize even quite improbable structures in a way that restricted beam searches simply cannot do.

Do other considerations undermine our argument for top-down analysis? In Chomskian syntax, linguists typically do their derivations bottom-up. Does this indicate that the theory requires bottom-up analysis? First, let's see that this conclusion does not follow in any straightforward way. Merge is a function, and so, like any other function, it can be regarded as a set of pairs. For example, for any syntactic objects x, y, merge might be simply the function with pairs $\langle\langle x,y\rangle,\{x,y\}\rangle$. That representation of the function certainly does not indicate anything about how complex representations $\{x,y\}$ are built in temporally unfolding psychological processes. A *bottom-up* parser builds the two pieces x and y first and then constructs the indication $\{\cdot\}$ that they are sisters in one structure. A *top-down* parser constructs the indication of a complex $\{\cdot\}$ first and then puts x and y into it. Note that the terms 'bottom-up' and 'top-down' here refer to the temporal order among the steps in constructing the complex representations $\{x,y\}$. Bottom-up parsers construct the parts first, then the indication of their relation to a larger phrase.[7] Does linguistic theory tell us which temporal order of these steps is correct? No. The sorts of evidence that linguists usually consider do not bear on the manner in which representations are constructed in temporally unfolding linguistic processes. Structural properties do have consequences for how they could possibly be constructed – parsers that construct the appropriate relations between pronounced forms and syntactic structures have certain memory requirements, etc. – but top-downness is not an issue decided by the structures themselves. This is famously observed by Chomsky (1959:p137n1), for example.

Uriagereka (2012:§3) argues for the opposite view, claiming that empirically secure aspects of linguistic theory indicate that representations are constructed bottom-up. Top-down models require extra, unmotivated "computational coding", he says, and then "any computable hack becomes as good as any other". Bottom-up models, on the other hand, require no more than "adding a bracket to an association". Why does he think that grammar as we currently understand it indicates that the 'bracket' is introduced after the parts? It is because he thinks that being first-merged in a bottom-up derivation is a natural property in our grammars, while the property of being last-merged in a top-down derivation is a hack that would require a whole different formal system:

> Linguistically speaking, the notion of a first-bottom-up-association (e.g. a first Merge) makes very good sense, and it appears in various domains, typically defining 'complements'... One could, in point of fact, express the core dependency just alluded to in terms that actually do not track the bottom-up character of computations. For example, rather than [saying] the item first merged to a verb determines its internal aspect, we may say that it is a sister to the verb that does. Now what is the notion 'sister', but one of those computational codings that tracks some helpful compilation?... by the usual compile-and-label hack we can turn that [i.e. the contribution of first-merged elements] into the notion *whatever-is-sister-to-a-head-lastly*... What is last can be defined too, no doubt, but it presupposes an

[7] The indication of the larger phrase, the common membership of the parts in a set or whatever, shown here as $\{\cdot\}$, could be links to a common element, as in the graphs of sets shown in Barwise and Moss (1996) and discussed in Gärtner (2002), but various other similar implementations of the relation can be imagined, and many ideas can be found in the formal parsing literature.

> entire formal system, and moreover some *limit* within which the 'lastness' can be bounded. (p.15, orig.emph.)

The key to this argument is the last line, and no argument is provided for it. In fact, the very idea seems to rest on a confusion between the specification of relations between syntactic objects and algorithmic theories about how those relations can be computed. Whether arguments and values of the merge function are constructed top-down or bottom-up, it is the same function! Top-down and bottom-up approaches do not need to differ *at all* on the function being computed; they can both compute exactly the same structures. The difference between top-down and bottom-up is not structural but temporal, and the theory of language structure says nothing whatever about the temporal order of steps in constructing representations. We want our algorithmic theory to be as simple and insightful as possible – not a 'hack' – but both bottom-up and top-down parsing models require resources (not specified by linguistic theory) in order to appropriately match analyses with perceived input, and any argument that one is simpler and more insightful, or more efficient, is going to have to be careful about the formal details. I cannot see any reason at all to suppose that any persuasive argument of that kind can be constructed. (Indeed, the point of this paper is that the facts seem to call for a top-down memoizing parser, which, as noted in the discussion of Graham et al. (1980) below, is really a mix top-down and bottom-up operations. Anyone who knows a more empirically adequate or more elegant approach should present it!)

One notable thing about such debates in the informal literature is that they often presume a kind of connection between the derivations of the theoretical linguist, on the one hand, and the temporal psychological processes, on the other, which is entirely inappropriate. It is often noted that the really revolutionary idea in Chomsky's proposals in the 1950's and since is the rejection of empiricist constraints on theory formation (Newmeyer 1986). A grammar is not a description of a corpus, nor is it the description of a set of behaviors, nor is it a description of the temporally unfolding psychological processes that create structures. It is not an algorithm for structure building that is implemented in neurophysiology. Chomskian linguistics has succeeded to such a remarkable extent by factoring structural properties away from the many other influences on linguistic behavior and mental processes, influences that depend on many different sorts of things (frequencies of lexical items, perceived social context, emotional state, competition for memory, and many other things). The rather narrow determinants of linguistic structure can be factored away from the much broader sorts of factors that influence what we will do in one or another linguistic performance. One respect in which the grammar is hidden from direct observation, like the epicenter of an earthquake, is that the determinants of grammatical structure are just one aspect of the complex causal nexus in any language performance.

To make these points a little more tangible, consider again the top-down parsing algorithm proposed by Stabler (2013). That parser uses MGs.[8] First, we can notice that the Stabler (2013) parser conforms to the structure of the grammar in proposing that the causal determinants of structure building are distinct from aspects of comprehension that are determined by different sorts of properties. That factoring of influences is needed in theories of parsing too. The actual performance of the model – which parses are constructed – is

[8] It is important to remember MGs are a simple formalization of Chomskian proposals that do not pretend to be fully adequate, as discussed by Stabler (2010) for example. But they are close enough to live proposals to illustrate a couple of issues relevant here.

explicitly mediated by other factors that are only roughly indicated in the paper. But a second point to note, more directly relevant to the present issue, is that the Stabler (2013) parser is initially defined in a way that conflates left-to-right order with derivation construction. The steps of the parser are described as fixing both hierarchical structure and left-to-right order at once. This is done to make the specification of the parsing algorithm as easy to understand as possible. In a separate appendix of that paper (Appendix B), the definition of the derivation and the definition of left-to-right pronounced order are disentangled yielding a simpler conception, but one which takes more steps to relate to algorithmic analysis. Linguistic theory properly factors these two aspects of linguistic structure, but in processing they must unfold together in an appropriate way.[9]

2 Exhaustive top-down memoizing parsing

The median isn't the message. (Gould 1985)

Goodbye Pareto principle, Hello long tail. (Brynjolfsson et al. 2011)

The most natural way to avoid the complexity problems of standard top-down parsing methods is to remember ('memoize') the steps of each analysis so that they need not be repeated. With this approach, even exhaustive searches for derivations can be tractable, if we separate the step of choosing a particular analysis from the calculation of all possibilities. With these dynamic programming approaches to parsing, the collection of a single parse is standardly postponed to a second step.

For context free grammars, one standard bottom-up memoizing parser is the CYK parser (Aho and Ullman 1972:§4.2.1). That parser can waste time building structures that could not be part of any successful parse. The Earley parser is a memoizing parser that constructs top-down derivations to determine what constituents could be part of a successful parse, restricting parse construction to those (Aho and Ullman 1972:§4.2.2). Both methods are exhaustive, finding all analyses. And both of these methods can be extended to MCS grammars, and to MGs in particular (Harkema 2001b; Kallmeyer 2010; Angelov 2011; Ljunglöf 2012). Further work is needed here to define these strategies in the simplest and most illuminating way, but their basic properties are well-understood.

Graham et al. (1980) observe that Earley parsing and CYK parsing are less different than their respective 'top-down' and 'bottom-up' labels might suggest. Top-down Earley parsing has bottom-up parse construction steps, but they are restricted to constituents that have been predicted. Relaxing the requirement that bottom-up steps be predicted makes parsing less efficient but does not affect the correctness and completeness of the method. So really, Earley's method and similar top-down memoizing parsers are mixtures of top-down and bottom-up steps. For incremental processing, the explicit top-down calculation of predictions has the great advantage that it defines connected, interpretable grammatical structures. But those connected structures are not needed for the determination of whether the input is derivable. One or more of them can be built for interpretation ('spellout to LF'),

[9] A third kind of possible difference between grammar and the parsing algorithm is proposed in Stabler (2012): we need not (and I think: should not) assume that every basic step in the derivation corresponds to a basic step in the computation of a parse. A similar idea is proposed by Fodor (2001) with the addition of 'treelets' to the basic building blocks of the language parser and learner.

but this is something the recognition process does not require. The recognition process uses the (exhaustively) predictive steps to avoid inefficiency, so that it can compute all possible analyses while avoiding wasted effort on analyses which are impossible in the sense of being inconsistent with the input.

Linguists might think that top-down memoizing parsers are engineering hacks, but that would be a mistake. The CFG parsers proposed by Earley (1968) and Graham et al. (1980) and others are simple and well-tailored to the requirements of the grammar. The extensions to MCS parsing (Harkema 2001b; Kallmeyer 2010; Angelov 2011; Ljunglöf 2012) are similarly designed to exactly fit their grammars. The goal in their design is to do exactly what the grammar requires, neither more nor less. So if the grammars of human languages are MCS, these parsers provide a good starting place. These parsers have been neglected, though, for reasons that now seem unsound. In fact, they are rarely explicitly considered in the psycholinguistic literature, though some informal suggestions come close. Briefly reflecting a little on this situation, in the next section, it appears that top-down memoizing models are easily compatible with the best established phenomena, and we suggest a number of hypotheses about how the facts of human performance could be treated, leaving more careful assessment to future work.

3 A *prima facie* defense of exhaustive memoization

It may seem strange that computing one parse at a time is intractable or even non-terminating, while constructing all parses in parallel can be efficient. But that is a slightly misleading way to describe the situation. While it is true that exhaustive parsing methods can be more efficient because, by remembering each step, no step needs to be repeated, this efficiency gain comes with a change in what is usually meant by ‘parsing’. ‘Parsing’ no longer means computing distinct representations of each possible parse. All possible steps of all parses are computed, and the construction of a particular one (e.g. for interpretation) is left to a second step. This is a concern for incremental models. Another concern is that, for standard exhaustive parsing methods, memory requirements grow without bound even for purely right-branching and left-branching structures. Let’s briefly consider whether these problems and related matters provide grounds for rejecting all-paths-at-once top-down parsing methods.

Memory bounds and center embedding. Center embedded structures seem to be especially difficult for people to analyze, across languages. A number of different possible explanations have been considered for this fact. Chomsky and Miller (1963) consider several ideas. Kimball (1973) makes the simple suggestion that the human parser cannot parse the constituents of more than two incomplete sentences at once, and roughly similar ideas are considered by Stabler (1994) and Lewis (1995). Gibson and Thomas (1997) suggest that a better account comes from counting not just the number of incomplete structural relations, but specifically the number of incomplete thematic relations. Gibson (1998) and Warren and Gibson (2002) argue that it is not the thematic requirements that matter so much as the givenness status of the embedded subjects; the new discourse referents impose a burden.

The particular concern about top-down memoizing parsers is that, as standardly implemented, their memory requirements increase without bound even for purely right-branching or purely left-branching structures. In purely right-branching structures, for example, the only constituents other than the single pronounced elements are the suffixes of the input; that is, all the categories are completed at the right edge, after the last word has been parsed. The 'complete' step in Earley-like parsing methods will register all these completions at the last step, referring arbitrarily far back into the analysis to find the predicted element that has been completed. It is possible to complete these elements earlier (e.g. when their last sub-constituent has been predicted), but with the increasingly fine-grained diagnosis of the difficulty of center-embeddings reviewed just above, it is not clear that the psychological model requires that.

Garden paths and reanalysis due to attention and spellout. As noted at the beginning of this section, Earley-like parsing methods separate the recognition step, finding a representation of all possible derivations, from what might be called the parsing or spell-out step: collecting a particular parse or particular parses from the set of possible derivations. In engineering applications, tree-collection is typically performed at the end of the parse, beginning at the right edge, so that dead ends, useless parser steps, can be avoided. But it is of course possible to define parse collection methods that are (incomplete but) left-to-right and incremental. In probabilistic models, Viterbi-like methods can find the most probable parse at each step, left-to-right, and clearly similar methods are possible using other sorts of salience or relevance.

The existence of garden path difficulties in processing some sentences is often thought to support the idea that one structure is built at a time (Fodor and Inoue 2000; Sturt et al. 1999; Frazier 1978; Bever 1970):

The horse raced past the barn fell.
Tom told the children the story scared a riddle.

However, Grodner et al. (2003) and others point out that these phenomena do not require special repair mechanisms, but can use standard derivation-building mechanisms to collect alternative parses.

Eye-movements seem to provide a much better, more sensitive window on garden paths in reading than earlier self-paced reading studies. In a review of the eye-tracking literature, Rayner and Pollatsek (2006) carefully conclude "readers either do not consider multiple syntactic analyses in parallel (Frazier 1978, 1987), or if they do, competition between these analyses does not disrupt processing (Van Gompel et al. 2001, 2005)." The latter idea is compatible with the top-down memoizing approach: readers may attend to one salient reading, and have difficulty switching to an alternative. The whole range of proposals about how and why that happens seem compatible with the view that the pieces of all analyses are ready to assemble.

More recent ERP studies show that prosodic cues are relevant in understanding spoken garden paths, and can confuse listeners when they are inappropriate Bögels et al. (2013); Pauker et al. (2011). This too is compatible with the hypothesis that prosody is guiding the use of standard derivation-collection methods defined over the whole collection of possible constituents found by the memoizing parser.

If the parser makes all pieces of all analyses available, in principle, then garden-path and incremental interpretation effects can be be due attention effects. Aiming to understand what is being heard, the subject usually attends to one analysis, sometimes noticing ambiguities, and sometimes failing to notice them in garden path constructions.

Two other reasons to favor exhaustive parsing. There are two places in psycholinguistic theory where exhaustive parsing has already been the norm. First, Hale (2001) noticed that entropy reduction can predict online complexity, and subsequent studies (Levy 2008, 2005; Staub 2010; Clifton and Staub 2011) confirm this result. Entropy is a global measure though, and so these effects would be surprising unless there is exhaustive parsing.[10] The second place in psycholinguistic theory where exhaustive parsing is ubiquitous is in learning models. Johnson (2013) and Hunter and Dyer (2013) consider statistical models which require exhaustive parsing of the input corpus in order to adjust the probabilities or weights of constituents to increase probability, or to move towards a feature-based maximum entropy. Clark and Yoshinaka (2012) and Yoshinaka (2011), on the other hand, aim to derive the constituents and rules directly from the input corpus, and so factor strings in all possible ways to find rules that fit the corpus at each point. If any of these ideas is on the right track, it seems that the language user must be able to parse exhaustively. It is very unclear how anything other than an exhaustive parser could explain entropy reduction effects in parsing or maximum entropy calculations in learning.

4 Linguistics as cognitive psychology

Given the range and importance of linguistic activities in human culture, it is no surprise that linguistics is a very broad diverse field. One recent and prominent tradition in the field aims to use scientific methods to obtain a clear and explanatory analysis of the core mechanisms of language. The focus of this effort is the grammar, since grammar is the definition of the common structural properties of expressions that seem to underpin a very wide range of linguistic activities, from reading essays in linguistics to hearing the poetic flourishes of Shakespeare on stage, from the ordinary conversations among ASL speakers to the silent musings we might have on a long walk, thinking about what to say to a friend. As Bever (1970) notes, this common structure in the vast range of linguistic activities is like a hidden epicenter behind the enormous impact of language on all of our activities.

There remains serious skepticism in the field about whether a scientific account of this sort is even possible. Is there any more reason to believe that there could be a scientific linguistics defining fixed limits on variation, specifying the mechanisms responsible for them, than there is to believe that there could be a generative biology defining fixed limits on organismic, phylogenetic variation and the responsible mechanisms? I think the prospects for a scientific linguistics are very much better, because in language we find a restricted range of structural options recurring over and over, derivable in each case as a function of a very restricted range of properties of basic elements. A very narrow range of basic lexical variation seems to be responsible for a very large range of significant properties of structural

[10] Levy (2005:p22) uses an exhaustive parser to calculate entropy, but slips over one of their problems when he suggests that these methods analyze sentences "using only a modest quantity of computational resources." In fact, these methods seem to require far too many steps on easy sentences. Aho and Ullman (1972:p.314) say of CF memoizing parsers that "n^3 time is too much to allow for parsing," noting also that these method need an amount of space proportional to the square of the input length. For MCS languages, memoizing parsers require even more resources, more than a human model can plausibly provide. This problem is left for future work!

complexes. The mechanisms defining this relation are what we aim to specify in grammar.

Of course, particular human languages vary enormously, but there are fairly secure claims about the kinds of structures found in all of them. This paper begins with one such claim, the hypothesis that human languages are (weakly and strongly) mildly context sensitive. And though there is notable individual variation in linguistic abilities, there are fairly secure and uncontroversial claims about how those structures are computed in recognizing ordinary fluent speech. This paper adopts one well-supported claim of this kind too, namely that human language analysis is typically incremental. So it is plausible that any reasonable psychological model requires an incremental parser for mildly context sensitive languages. There is a wide and growing range of proposals (Kallmeyer and Maier 2009; Angelov 2009; Shen et al. 2008; Dubey et al. 2006; Shieber and Johnson 1994:and many others), but it is still not clear how such a parser can accommodate even the most basic facts. This paper defends a rather simple kind of proposal: a predictive dynamic programming approach can use the mechanisms of grammar transparently; it can be efficient by memoizing steps, so they are not repeated; incremental interpretation can then be an attention effect, and reanalysis can be the standard construction of grammatical alternatives from remembered parts. Careful empirical assessment is required to tease the predictions of this perspective apart from close alternatives, but it does seem to be worth considering.

References

Aho, Alfred V., and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation, and Compiling. Volume 1: Parsing*. Englewood Cliffs, New Jersey: Prentice-Hall.

Altmann, Gerry T. M., and Yuki Kamide. 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73:247–264.

Angelov, Krasimir. 2009. Incremental parsing with parallel multiple context-free grammars. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL*, 69–76.

Angelov, Krasimir. 2011. The mechanics of the grammatical framework. Doctoral Dissertation, Göteborg University.

Barwise, Jon, and Lawrence Moss. 1996. *Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena*. Lecture Notes Number 60. Stanford University: CSLI Publications.

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In *Cognition and the Development of Language*, ed. J.R. Hayes. NY: Wiley.

Bögels, Sara, Herbert Schriefers, Wietske Vonk, Dorothee J. Chwilla, and Roel Kerkhofs. 2013. Processing consequences of superfluous and missing prosodic breaks in auditory sentence comprehension. *Neuropsychologia* 51:2715–2728.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Duncan Simester. 2011. Goodbye Pareto principle, Hello long tail. *Management Science* 57:1373–1386. URL http://dx.doi.org/10.1287/mnsc.1110.1371.

Chambers, Craig G., Michael K. Tanenhaus, Kathleen M. Eberhard, Hana Filip, and Greg N. Carlson. 2002. Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language* 47:30–49.

Chomsky, Noam. 1959. On certain formal properties of grammars. *Information and Control* 2:137–167.

Chomsky, Noam, and George A. Miller. 1963. Introduction to the formal analysis of natural languages. In *Handbook of Mathematical Psychology, Volume II*, ed. R. Duncan Luce, Robert R. Bush, and Eugene Galanter, 269–321. NY: Wiley.

Church, Kenneth, and Ramesh Patil. 1982. How to put the block in the box on the table. *Computational Linguistics* 8:139–149.

Clark, Alexander, and Ryo Yoshinaka. 2012. Beyond semilinearity: Distributional learning of parallel multiple context-free grammars. In *Proceedings of the 11th International Conference on Grammatical Inference, ICGI 2012*. http://jmlr.org/proceedings/papers/v21/clark12a/clark12a.pdf.

Clifton, Charles, and Adrian Staub. 2011. Syntactic influences on eye movements in reading. In *The Oxford Handbook of Eye Movements*, ed. S.P. Liversedge, Iain D. Gilchrist, and Stefan Everling, 895–909. Oxford: Oxford University Press.

Clifton, Charles, Adrian Staub, and Keith Rayner. 2006. Eye movements in reading words and sentences. In *Eye Movements: A Window on Mind and Brain*, ed. R. P. G. von Gompel, M. H. Fischer, W. S. Murray, and R. C. Hill. Oxford: Elsevier.

Dubey, Amit, Frank Keller, and Patrick Sturt. 2006. Integrating syntactic priming into an incremental probabilistic parser, with an application to psycholinguistic modeling. In *Proceedings of the Association for Computational Linguistics, COLING/ACL 2006*, 417–424. Association of Computational Linguistics.

Earley, J. 1968. An efficient context-free parsing algorithm. Doctoral Dissertation, Carnegie-Mellon University.

Fodor, Janet Dean. 2001. Setting syntactic parameters. In *The Handbook of Contemporary Syntactic Theory*, ed. Mark Baltin and Chris Collins, 730–767. Oxford: Blackwell.

Fodor, Janet Dean, and Atsou Inoue. 2000. Garden path reanalysis: Attach (anyway) and revision as last resort. In *Cross-linguistic Perspectives on Language Processing*, ed. M. De Vincenzi and V. Lombardo, 21–61. Boston: Kluwer.

Frazier, Lyn. 1978. On comprehending sentences: Syntactic parsing strategies. Doctoral Dissertation, University of Massachusetts, Amherst.

Frazier, Lyn. 1987. Sentence processing: a tutorial review. In *Attention and Performance XII: The Psychology of Reading*, ed. M. Coltheart, 559–586. Hillsdale, New Jersey: Lawrence Erlbaum.

Frazier, Lyn. 1999. *On Sentence Interpretation*. Dordrecht: Kluwer.

Gärtner, Hans-Martin. 2002. *Generalized Transformations and Beyond: Reflections on Minimalist Syntax*. Berlin: Akademie-Verlag.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68:1–76.

Gibson, Edward, and James Thomas. 1997. Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical. Massachusetts Institute of Technology.

Gould, Stephen Jay. 1985. The median isn't the message. *Discover* 6:40–42.

Graham, Susan L., Michael A. Harrison, and Walter L. Ruzzo. 1980. An improved context free recognizer. *Association for Computing Machinery Transactions on Programming Languages and Systems* 2:415–462. http://dl.acm.org/citation.cfm?id=357112.

Grodner, Daniel, Edward Gibson, Vered Argaman, and Maria Babyonshev. 2003. Against repair-based reanalysis in sentence comprehension. *Journal of Psycholinguistic Research* 32:141–166.

Hale, John. 2001. Uncertainty about the rest of the sentence. In *Proceedings of the Second Meeting of the North American Chapter of the Asssociation for Computational Linguistics*, N01–1021.

Harkema, Henk. 2001a. A characterization of minimalist languages. In *Logical Aspects of Computational Linguistics*, ed. P. de Groote, G. Morrill, and C. Retoré, LNCS 2099, 193–211. NY: Springer.

Harkema, Henk. 2001b. Parsing minimalist languages. Doctoral Dissertation, University of California, Los Angeles.

Hunter, Tim, and Chris Dyer. 2013. Distributions on minimalist grammar derivations. In *Proceedings, Mathematics of Language*. http://wwwhomes.uni-bielefeld.de/mol/mol13/3.pdf.

Jäger, Gerhard, and James Rogers. 2012. Formal language theory: Refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society B* 367:1956–1970.

Johnson, Mark. 2013. Language acquisition as statistical inference. In *19th International Congress of Linguists, CIL19*.

Joshi, Aravind K. 1985. How much context-sensitivity is necessary for characterizing structural descriptions. In *Natural Language Processing: Theoretical, Computational and Psychological Perspectives*, ed. D. Dowty, L. Karttunen, and A. Zwicky, 206–250. NY: Cambridge University Press.

Jurafsky, Daniel. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20:137–194.

Kallmeyer, Laura. 2010. *Parsing Beyond Context-Free Grammars*. NY: Springer.

Kallmeyer, Laura, and Wolfgang Maier. 2009. An incremental Earley parser for simple range concatenation grammar. In *Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09)*, 61–64. Association for Computational Linguistics. URL http://alpage.inria.fr/iwpt09/.

Kimball, John. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2:15–47.

Kobele, Gregory M. 2006. Generating copies: An investigation into structural identity in language and grammar. Doctoral Dissertation, UCLA.

Levy, Roger. 2005. Probabilistic models of word order and syntactic discontinuity. Doctoral Dissertation, Stanford University, Stanford, California.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106:1126–1177.

Lewis, Richard L. 1995. A theory of grammatical but unacceptable embeddings. Technical report, Princeton University.

Ljunglöf, Peter. 2012. Practical parsing of parallel multiple context-free grammars. In *TAG+11, 11th International Workshop on Tree Adjoining Grammar and Related Formalisms*.

Manaster-Ramer, Alexis. 1986. Copying in natural languages, context freeness, and queue grammars. In *Proceedings of the 1986 Meeting of the Association for Computational Linguistics*.

Marslen-Wilson, William. 1975. Sentence perception as an interactive parallel process. *Science* 189:226–228.

Michaelis, Jens. 2001. Transforming linear context free rewriting systems into minimalist grammars. In *Logical Aspects of Computational Linguistics*, ed. Philippe de Groote, Glyn Morrill, and Christian Retoré, Lecture Notes in Artificial Intelligence, No. 2099, 228–244. NY: Springer.

Michaelis, Jens, and Marcus Kracht. 1997. Semilinearity as a syntactic invariant. In *Logical Aspects of Computational Linguistics*, ed. Christian Retoré, 37–40. NY: Springer-Verlag (Lecture Notes in Computer Science 1328).

Newmeyer, Frederick J. 1986. Has there been a 'Chomskyan revolution' in linguistics? *Language* 62:1–18.

Norris, Dennis, James M. McQueen, and Anne Cutler. 2000. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23:299–325. URL http://dx.doi.org/10.1017/S0140525X00003241.

Pauker, Efrat, Inbal Itzhak, Shari R. Baum, and Karsten Steinhauer. 2011. Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *Journal of Cognitive Neuroscience* 23:2731–2751.

Peters, P. Stanley, and R. W. Ritchie. 1973. On the generative power of transformational grammar. *Information Sciences* 6:49–83.

Pryor, James. 2005. What's so bad about living in the matrix? In *Philosophers Explore the Matrix*, ed. Christopher Grau, 49–61. NY: Oxford University Press.

Rambow, Owen. 1994. Formal and computational aspects of natural language syntax. Doctoral Dissertation, University of Pennsylvania. Computer and Information Science Technical report MS-CIS-94-52 (LINC LAB 278).

Rayner, Keith, and Alexander Pollatsek. 2006. Eye-movement control in reading. In *Handbook of Psycholinguistics, Second Edition*, ed. Matthew J. Traxler and Morton Ann Gernsbacher, 613–375. NY: Academic Press.

Roark, Brian. 2001. Probabilistic top-down parsing and language modeling. *Computational Linguistics* 27:249–276.

Roark, Brian. 2004. Robust garden path parsing. *Natural Language Engineering* 10:1–24.

Seki, Hiroyuki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami. 1991. On multiple context-free grammars. *Theoretical Computer Science* 88:191–229.

Shen, Libin, Lucas Champollion, and Aravind K. Joshi. 2008. LTAG-spinal and the Treebank: A new resource for incremental, dependency and semantic parsing. *Language Resources and Evaluation* 42:1–19.

Shieber, Stuart, and Mark Johnson. 1994. Variations on incremental interpretation. *Journal of Psycholinguistic Research* 22:287–318.

Stabler, Edward P. 1994. The finite connectivity of linguistic structure. In *Perspectives on Sentence Processing*, ed. Charles Clifton, Lyn Frazier, and Keith Rayner, 245–266. Hillsdale, New Jersey: Lawrence Erlbaum.

Stabler, Edward P. 2010. Computational perspectives on minimalism. In *Oxford Handbook of Minimalism*, ed. Cedric Boeckx, 617–641. Oxford: Oxford University Press.

Stabler, Edward P. 2012. Recursion in grammar and performance. Forthcoming. Presented at *Recursion: Structural Complexity in Language and Cognition*, University of Massachusetts, Amherst. http://www.linguistics.ucla.edu/people/stabler/Stabler10-Recurs.pdf.

Stabler, Edward P. 2013. Two models of minimalist, incremental syntactic analysis. *Topics in Cognitive Science* 5:611–633. http://www.linguistics.ucla.edu/people/stabler/Stabler12-2models.pdf.

Staub, Adrian. 2010. Eye movements and processing difficulty in object relative clauses. *Cognition* 116:71–86.

Staub, Adrian. 2011. Word recognition and syntactic attachment in reading: Evidence for a staged architecture. *Journal of Experimental Psychology: General* 140:407–433.

Sturt, Patrick. 2004. Incrementality in syntactic processing: Computational models and experimental evidence. In *Proceedings of the ACL Workshop on Incremental Parsing*, ed. Frank Keller, Stephen Clark, Matthew Crocker, and Mark Steedman, 66. Association for Computational Linguistics.

Sturt, Patrick, Martin Pickering, and Matthew W. Crocker. 1999. Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language* 40:136–150.

Tanenhaus, Michael K, and Sarah Brown-Schmidt. 2008. Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:1105–1122.

Tanenhaus, Michael K., James S. Magnuson, Bob McMurray, and Richard N. Aslin. 2000. No compelling evidence against feedback in spoken word recognition. *Behavioral and Brain Sciences* 23:348–349.

Uriagereka, Juan. 2012. *Spell-out and the Minimalist Program*. NY: Oxford University Press.

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

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

VanWagenen, Sarah, Jonathan Brennan, and Edward P. Stabler. 2011. Evaluating parsing strategies in sentence processing. In *Poster, CUNY Sentence Processing Conference*.

VanWagenen, Sarah, Jonathan Brennan, and Edward P. Stabler. 2014. Evaluating parsing strategies in sentence processing. In *this volume*.

Warren, T., and E. Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85:79–112.

Weir, David. 1988. Characterizing mildly context-sensitive grammar formalisms. Doctoral Dissertation, University of Pennsylvania, Philadelphia.

Yoshinaka, Ryo. 2011. Efficient learning of multiple context-free languages with multidimensional substitutability from positive data. *Theoretical Computer Science* 412:1821–1831.

Affiliation

Edward P. Stabler
Linguistics Department
UCLA
stabler@ucla.edu