Postscript

What has become of formal grammars in linguistics and psycholinguistics?

The aim of this postscript cannot be to review the theory and language science applications of formal languages and automata, as developed since the mid 1970s. That would require more than a three-volume work. I will, rather, touch upon just a few developments that seem to me of special relevance to linguists and psycholinguists. I will do this under the three main headings I used in *Formal Grammars.*

Formal languages and automata

Of special linguistic relevance has been the construction of tree grammars and tree automata. The original grammar types in the Chomsky hierarchy, as well as the corresponding automata were string handling devices. Their inputs and outputs were strings of symbols. Their structural descriptions consisted of the derivation or recognition trees as they emerged in the stepwise application of the rules. The newly introduced tree grammars and tree automata operate on trees, not on strings. In that sense they are operations on structural descriptions. A tree grammar generates a set of trees. The 'frontier' of a tree is its bottom string of symbols, which consists of (at least one) terminal and/or non-terminal nodes. The tree set generated by a tree grammar is the set of completed' trees derived from one or more special S-rooted initial trees. A tree is completed if its frontier consists of terminal elements only. The language generated by this tree grammar is the 'yield' of this tree set, i.e., the set of its (terminal) frontiers.

The strong generative power of the grammar is the set of terminal trees generated for this language. The theory of tree automata and grammars originated from Büchi (1960). A recent overview is presented in Comon et al. (2007).

The generative power of types of tree grammars does not simply match the power of types in the Chomsky hierarchy. An interesting equivalence holds between context-free languages and the languages generated by so-called regular tree grammars (cf. Gékseg & Steinby 1997). More generally, however, the power of tree grammars (and automata) straddles the power levels in the Chomsky hierarchy, which has interesting applications in linguistic theory (see below).

Ellis (1971) introduced the notion of probabilistic tree grammars and automata, which generate/accept probabilistic languages. He showed that context-free probabilistic languages (as defined in *Formal Grammars* II, 3.4) can be fully characterized by probabilistic tree automata.

Linguistic applications

An appropriate level of generative power

A perennial issue in formal linguistics has been the characterization of the 'right' level of grammatical power for natural language grammars. A major motivation for Chomsky's original work on formal grammars had been to show that finite state automata or regular grammars cannot characterize natural languages. Here the recursive self-embedding property of natural languages transcended the capacity of this type of system. Context-free grammars faired a lot better, but reached their limits in dealing with crossed and other long-distance dependencies. This was all comprehensively reviewed in *Formal Grammars* II. Initially, the move to transformational grammars seemed to be a promising one for handling such problems. However, Peters & Ritchie's proof (1973, see *FG* II, chapter 5) that the then most advanced transformational grammar, Chomsky's *Aspects* model, has the generative capacity of a Turing machine (see *FG* II, chapter 5), showed that simple or 'natural' solutions were not yet around. In his interview with Huybregts and van Riemsdijk, Chomsky (1982, p. 15) remarked:

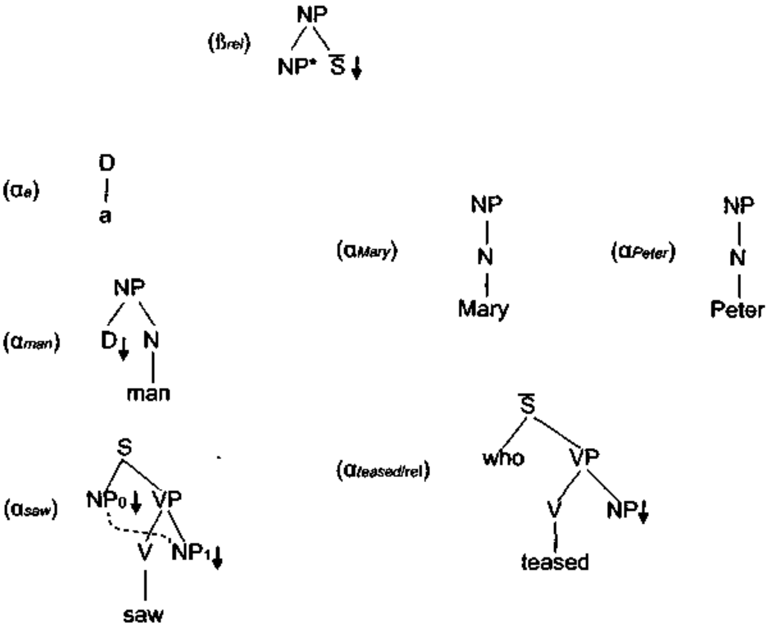
> "The systems that capture other [than the context-free WL] properties of language, for example transformational grammar, hold no interest for mathematics. But I do not think that that is a necessary truth. It could turn out that there would be richer or more appropriate mathematical ideas that would capture other, maybe deeper properties of language than context-free grammars do. In that case you have another branch of applied mathematics which might have linguistic consequences. That could be exciting" '

Such exciting formalisms were, then, about to emerge. There exists now a class of equivalent grammars, called 'mildly context-sensitive grammars' (MCSGs), among them linear indexed grammars, head grammars, combinatory categorial grammars

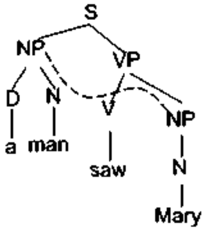
---

1. I thank Aravind Joshi for this reference.

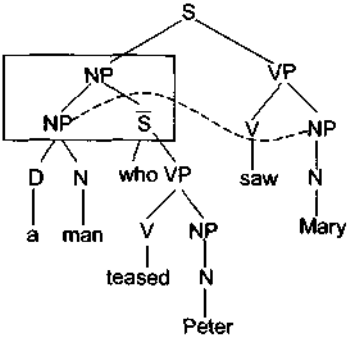
and tree adjoining grammars, that seem to have just the right level of generative capacity to overcome the linguistic limitations of context-free grammars without over-generating' to the level of context-sensitive grammars, or worse, of recursively enumerable languages. Let us, for a moment, consider the case of the tree adjoining grammars (TAG) conceived of by Joshi and his co-workers (cf. Joshi & Schabes 1997 and further references there). A TAG has a set of 'elementary trees', consisting of 'initial trees' and auxiliary trees'. In the following example seven such elementary trees from a TAG are exemplified (adapted from Joshi & Schabes, p. 75):



New trees are generated from the elementary trees by two operations, *substitution* and *adjoining.* Substitution is the operation by which the root of an initial tree (marked a above) substitutes for a node in the frontier of some tree which is marked for substitution (indicated by ↓ on that node). Substitution is allowed when certain constraints are met. In this simple example the only constraint is that the node labels of the marked node and the replacing root node match. (In the full grammar a Boolean comparison of feature sets is to be performed). So, for instance, the NP root node of the *Mary* tree can substitute for the NP$_1$↓ node of the saw tree. The following tree is entirely derived by substitution, using four of the initial trees above. The order of substitution is irrelevant. The terminal frontier of the tree is the sentence *a man saw Mary.*



Because a TAG has only a finite set of initial trees (to be motivated below), substitution only allows for the derivation of a finite set of derived trees. Recursion in a TAG is handled by the other operation, adjoining. It requires the insertion of an auxiliary tree (marked (3 above). An auxiliary tree has some leaf node (marked *) that is identical to its root node. That is the case for the β*rel* tree above, which has a root node NP and a leaf node NP* in its frontier. This tree can be (does not *have* to be) inserted into the above derived tree by detaching the NP tree dominating *a man,* attaching the *auxiliary* tree to the 'freed' NP node and finally attaching the detached a man tree to the starred NP node of the inserted auxiliary tree. The result is the derived tree below (with the inserted tree indicated by a square frame).



Substituting the S node by the *teased* tree and substituting the *Peter* tree in its NP node, results in a tree dominating the sentence *a man who teased Peter saw Mary.* Repeated adjunction of the auxiliary tree generates sentences such as *the man who teased Mary who teased Peter saw Mary,* etc.

This is just enough example[2] to notice that syntactic dependences (such as the c-command relation between NP_0 and NP_1 in the initial *saw* tree above, marked by a dotted line both there and in the two following derived trees) can be defined

---

[2] without any linguistic pretensions.

within the elementary trees. Recursion is exclusively realized by adjunction. Adjunction can move the elements between which a dependency relation exists arbitrarily far apart, thus accounting for long-distance dependences. TAGs elegantly separate the syntactic functions of dependency and recursion.

Like the other MCSGs, TAGs can handle crossed dependences, such as in Dutch complement constructions (e.g. ... *dat Jan Piet Marie zag laten zwemmen - that John Peter Mary saw let swim,* where John sees, Peter lets and Mary swims) (Stabler, 2004), which are problematic for a context-free account, and they can handle other similar phenomena. Their generative power is 'slightly' more than context-free, but less than context-sensitive. TAG grammars have equivalent accepting automata, called embedded push-down automata' (EPDA) and a parser has been defined which, incrementally, parses a string in the TAGs language 'from left to right', ultimately delivering the appropriate tree for the sentence. Parsing of TAGs, and probably more generally of MCSGs, is polynomial (in the worst case with $0(n^6)$ time and $0(n^4)$ space), hence hardly worse than for context-free grammars.

A closely related, but alternative tree adjoining grammar, 'Performance Grammar' (PG) has been introduced by Kempen & Harbusch (1998). Like in TAG, its trees are 'lexicalized' (see below) and it has essentially the same substitution operations as in TAG. However, there is no adjoining operation. Instead, the grammar has a topological or 'linearization component which can handle recursion and long distance dependences. Each (initial) lexical tree is combined with a linear array, which is a topology for the ordering of the lexical item's arguments. This topology goes back to the time-honored tradition in Germanic linguistics of distinguishing syntactic Vorfeld (forefield), Mittelfeld (midfield), and Nachfeld (endfield). The key recursive property of PG, which at the same time generates the appropriate linearization, is that constituents may move out of their 'own' array and receive a position in an array located at a higher level. The linearization operation is implemented as a finite state automaton. This mechanism provides PG with the generative power of MCSGs. At the same time, it elegantly handles the relatively free word order of German (Harbusch & Kempen 2002).

Lexicalization of grammars

Another major development since the 1970s is the 'lexicalization of grammars. The first context-free, and also transformational grammars, were rule systems fully abstracted from the ultimate lexical insertions. A rule such as VP —> V + NP applies whatever the ultimate lexical insertions (such as *saw* and *Mary).* The lexicon was the linguist's last concern, as it were. This approach was already challenged by the generative semanticists (see below), but the major later innovation was to characterize lexical items as syntactic structures, which would then 'bottom-up' interact or 'unify' to generate syntactic phrase structures. *Formal Grammars* (II, p. 94) already pleaded for such a move, but the idea caught the (psycho)linguistic community's imagination by Kaplan and Bresnan's (1982) work on their Lexical Functional Grammar (LFG). Still, the core notion had been around in categorial grammar since it was created by Ajdukiewicz (1935). This formalism is extensively discussed in *Formal Grammars* II 4.2). Each word has a specific syntactic category, which allows it to 'hook up' (or unify) with other words that have some 'hookable' category, more or less like Lego pieces.

The lexicalization of grammars was a major asset for psycholinguistic applications. Language users have a huge mental lexicon and they know the syntactic affordances of these lexical items. It is a natural idea that the listener, who recognizes one word after another, indeed, on-line, 'unifies' their syntax, thus incrementally building up the phrase structure of the incoming sentence. That has been Steedman's motivation all along in developing his Combinatory Categorial Grammar (CCG) (cf. Steedman 2000), which mediates both the on-line semantic interpretation and the interpretation/generation of prosody. My own book *Speaking* (1989) makes extensive use of LFG.

Joshi and Schabes' (1997) version of TAG is completely 'lexicalized'. This means, first, that all elementary trees have at least one lexical term in their frontier; it is their 'lexical anchor' (this is the case for six of the seven elementary trees above). It means, second, that all lexical items of the language figure as anchor in a finite number of elementary trees (at least once). This explains the earlier finite generative power of the substitution operation in TAG and also in PG. This type of lexicalization also provides a natural way of handling fixed expressions, such as *kick the bucket.* They have their own V-rooted elementary tree with a multiple lexical anchor. This is presently receiving interesting applications in language acquisition research (see below).

Semantics

Many such fixed expressions, in particular idioms, violate the 'principle of compositionality' (PC), which says that the meaning of the whole is a function of the meaning of the parts and the way they are syntactically combined. Five to ten percent of the words we speak are part of some fixed expression (Sprenger et al. 2006), which means that the principle could still have wide application in language use. In fact, it is basic to all formal semantics. The state of semantics discussed in *FG* II 3.3 was the then raging conflict between the generative semantics and the interpretative semantics approach. Both incorporated PC but in quite different ways. In the first approach syntax was itself semantic. Deep structures were as much syntactic as semantic representations (hence they were fully lexicalized) and transformations were, presumably, meaning preserving or 'paraphrastic'. Soon a 'prelexical syntax' developed, in which lexicalized subtrees could be transformationally replaced by other subtrees, in particular by unitary lexical items (for instance replacing 'cause to become not alive' by 'kill'). This obviously raised the power of generative semantics to Turing machine level.

In interpretative semantics the underlying deep structures were purely syntactic entities, but could receive semantic interpretation after lexical insertion. Here an autonomous' semantics had to be developed which would provide the 'logical form' associated with some deep or (later) surface syntactic structure. However, it soon turned out (see *FG* II, p. 109) that transformations could not preserve the semantic interpretation of quantifiers in deep or underlying structure *(John sings and dances* can be paraphrased as *John sings and John dances,* but *One boy sings and dances* cannot be paraphrased as *One boy sings and one boy dances).*

It is both beyond the aim of this postscript and beyond my competence to sketch the developments in formal semantics since these early beginnings. A few remarks should suffice. The generative semantics approach survived a rather dramatic history of upheavals, ultimately producing a broad and formal treatment of meaning in language as it is represented in the language user's mind. That has, in its later developments, largely been Pieter Seuren's achievement, now available as Seuren (2009). It was in generative semantics and in Harman (1970) that a truth-functional semantics was first introduced in generative linguistics. Seuren's (1969) 'operators' (quantifiers, modal, tense and other operators) were truth-functional operations on their arguments, the so-called nuclei. The nuclei are the elementary propositions, which can be negated, questioned, etc. Harman extended the operator approach to the nuclei themselves, defining the main verb as an operator on the other phrases as arguments. It is now commonplace to analyze linguistic expressions as function-argument structures, but that idea was totally absent in early interpretative semantics. Although the truth-functional approach is now basic to any formal semantics, the cognitive perspective which has always been essential in generative semantics, has made the latter a laboratory for studying the many other aspects of meaning involved in the listener's on-line interpretation of language. Among them are the fascinating complexities of presupposition, discourse, anaphora, metaphor and lexical meaning. Here, Seuren (2009) provides a rich source for psycholinguists.

The early autonomous approach was completely transformed under the influence of Montague's (1970) truth-functional approach to natural language semantics. It was in particular Barbara Partee who managed to fuse the Chomskyan and Montagovian traditions, using lambda extraction to handle variable binding (see her own wonderful account of these and later developments in Partee 1997). Basic to Montagues handling of compositionality is the 'rule-by-rule' correspondence between syntax and semantics. Each syntactic composition of 'smaller' or 'lower-level' syntactic entities goes with a semantic interpretation of the higher-level entity in terms of the semantic interpretations of the lower-level units. This homomorphism between syntax and semantics has found wider application, for instance in TAG semantics. There it is not the phrase-structural relations of the derived tree that receive semantic interpretation. Rather, each substitution or adjoining application goes 'synchronously' with a corresponding semantic interpretation. It is therefore the derivational history (represented in a 'derivation tree') that provides the step by step correspondence to semantic operations that generate the 'logical form' of the linguistic expression. Although Partee always intended the formal semantics developed in the Montagovian tradition to be a theory of meaning in the mind, it didn't really conquer the hearts of psycholinguists. The 'possible worlds' framework and its somewhat daunting formal rendering, rightly or wrongly, always remained somewhat unapproachable for the psycholinguist studying the *process* of 'on-line' semantic interpretation in the language user's mind.

A third major approach has been Jackendoff's (2002). Coming from the interpretative tradition, but doing away with Chomsky's syntactocentrism, he developed, in much detail, a semi-formal cognitive theory of grammar with three parallel generative components, a conceptual, a syntactic and phonological one. In *Speaking* (1989) I gratefully used a version of Jackendoff's conceptual component. The system is still 'interpretative' in that it handles semantic interpretation by means of correspondence rules' that hold between conceptual and syntactic structures (just as phonological interpretation is handled by correspondence rules between phonological and syntactic structures). Jackendoff handles a great variety of meaning aspects, which have obvious psycholinguistic applications.

Many of these meaning aspects have conversational impact. A major challenge is to sort out how semantics and pragmatics interact in conversational implicature and anaphora. Levinson (2000) advocated a strong pragmatic stance here, which is probably less amenable to formalization than, for instance, Seuren's more formal semantic approach to these matters.

Probabilistic grammars and linguistic intuitions.

When I wrote *Formal Grammars,* probabilistic grammars were generally avoided by linguists as 'not done'. "It must be recognized that the notion of 'probability of a sentence' is an entirely useless one, under any known interpretation of this term", Chomsky wrote (see *FL* II, p. 174). Mine was the only text for linguists around that treated them. The only linguistic example of a probabilistic context-free grammar I could find at the time was the one that Patrick Suppes, always averse to current dogma, had written for a child language corpus (see *FG* II 6.2). The next one I came across was the probabilistic CFG Wolfgang Klein published, as early as 1974, for handling a large corpus of untutored second language (German) acquisition data. Meanwhile, however, stochastic approaches to grammars, automata, parsers, inference devices, automatic translation have exploded. A landmark publication was Charniak (1993), in which a wide range of (in some cases still potential) linguistic applications of probabilistic grammars was treated. Since then, the computational analysis of ever larger natural language corpora has stimulated the further development and use of probabilistic tree grammars and automata, which Ellis (1970) had initiated. Shabes (1992), for instance, introduced probabilistic TAGs, with Resnick (1994) applying them to natural language parsing. For a more recent review of stochastic tree approaches to natural language processing, see Knight & Graehl (2005).

One issue addressed in *FL* II, chapter 1, was the status of linguistic intuitions. At that time, generative linguistics was, as an empirical science, largely intuition-based. The grammaticality judgment played the essential role in telling 'grammatical' from 'ungrammatical' strings. In *FL* III I argued that grammaticality judgments are the outcome of metalinguistic judgment, a psychological process whose workings were still largely in the dark. And worse, I could provide some empirical evidence for their alarming unreliability. I analyzed various causes of this unreliability and proposed a range of empirical procedures to improve on this empirical weakness in linguistic practice. It didn't help much. Nor did Bard et al.'s (1996) careful proposal to use easily applicable magnitude estimation. Many linguists still mark strings as ungrammatical (by '*'), without providing their empirical reasons for doing so. Their tacit assumption is that they are dealing with god-given grammaticality', not with human 'acceptability'.

Luckily, the use of large corpora has meanwhile reduced the importance of grammaticality judgments. But it also raised the new issue whether some consistent relation exists between grammaticality judgments and statistical corpora data. Bresnan's (2006) study of linguistic intuitions seems to show that there is, indeed, a close relation between naturalness/acceptability judgments and corpus frequency data. But that requires judgments to be made in the appropriate textual context. In one experiment she had subjects rate the 'naturalness' of two alternative sentential continuations of short texts from a natural language corpus. The two sentences both contained a dative verb construction, but differed in whether the construction was prepositional or double object (for instance: *because he brought* *the pony to my children* versus *because he brought my children the pony).* It turned out that the naturalness judgments were highly predictable from the syntactic probabilities in the corpus model. What about really 'ungrammatical' sentences? In a second experiment subjects judged dative constructions that linguists usually mark as ungrammatical, such as *the dealer pushes someone the pot.* The stochastic corpus model, however, predicted contexts in which such sentences would appear, even with higher probability than the 'grammatical' alternatives (such as *the dealer pushes the pot to someone).* Again naturalness judgments followed the corpus model, not the linguists' judgments. Bresnans conclusion was that grammaticality judgments reflect implicit knowledge of syntactic probabilities.

Others, however, observed systematic disagreement between judged grammaticality and corpus probability. There is even talk about a grammaticality-frequency gap'. Kempen and Harbusch (2005, 2008) compared available judgment data for various German word order patterns (German allows for six different orderings of subject and objects in double object sentences) to frequencies of occurrence in two text corpora. One surprising finding was that lower rated word orders never occurred in the corpus. (It was unlikely that corpus size was an important factor here.) Another was that similarly highly rated word order types turned out to have quite different corpus frequencies. Extensive analyses of these data led the authors to make specific claims about the grammaticality judgment process. Judgers will normally try to internally generate the target sentence. If it works, it will be judged (highly) 'grammatical'. If it doesn't work, the subject will generate a sentence with the same semantic gist and then judge its *similarity* to the target sentence. In this way, highly unlikely sentences can still (by similarity to likely sentences) be judged as (somewhat) grammatical although they (or their type) didn't make it into the corpus. In other words, this similarity factor would deserve careful control, in addition to all the other reliability undermining factors I discussed in *FL* III, chapter 1. The Kempen & Harbusch studies cannot be directly compared to Bresnan's, because the grammaticality judgments were made on sentences in isolation. One wonders in particular whether the grammaticality-frequency gap will also appear in her data when the same experimental sentences are judged in isolation.

My one original psychological contribution to the study of linguistic intuitions in *Formal Grammars* is a mathematical theory of syntactic relatedness intuitions {*FG* III, 27-65). We have, for instance, the strong intuition that in the sentence *John ordered a pizza* the syntactic relation between *a* and *pizza* is much stronger than between *John* and *a* or between *John* and *pizza.* The mathematical theory relates such cohesion intuitions to the structural descriptions grammars adduce to sentences. That makes it possible to test the descriptive adequacy of (different kinds of) grammars in an entirely new way. The initial empirical tests, reported in *FL*, showed that transformational dependency grammars excelled on this test (see Schils 1983 for more extensive data and analyses). The method has also been successfully applied by Fodor et al. (1980) to distinguish between alternative 'underlying' structural descriptions. Take the two sentences (1) *the captain persuaded the passengers to leave* and (2) *the captain expected the passengers to leave.* Here one would expect the syntactic cohesion between *captain* and *passengers* to be stronger in (1) than in (2). This is because in the former, but not in the latter, the two items are in the same underlying clause (the *captain persuaded the passengers* S). And this was indeed found in the rating experiment. Fodor et al. could then use this sensitive procedure to test whether causative verbs (as in *John killed Mary)* have an underlying structure like (2) (i.e., *John caused Mary to die),* where *John* and *Mary* do not share a clause, or rather the simple 'non-definitial' one *John killed Mary,* just like for *John liked Mary,* where they do. The rating results were crystal-clear: the latter was the case. There is no evidence for a 'definitial' underlying structure of causative verbs. It is my impression that cohesion judgments are more reliable, less vulnerable than grammaticality judgments. They are, moreover, *direct* tests of descriptive adequacy, as opposed to grammaticality judgments, which concern strings, not structures. In short, intuitions of syntactic cohesion should still be embraced by linguists.

Psycholinguistic applications

Incrementality

An essential feature of modern theories of speaking and speech comprehension is incrementality. Speakers work with quite restricted 'look-ahead' (Levelt 1989). And the evidence is overwhelming that listeners interpret speech largely 'on line' as it comes in. All relevant knowledge (phonetic, phonological, morphological, syntactic, semantic, pragmatic) is in no time applied to any next incoming signal. Interpretations are rarely (but not never!) delayed or revised. In the early 1970s this insight was not yet around. Tom Bever's 'garden path' sentence *the horse raced past the barn fell* was on everybody's mind. Incrementality severely restricts the nature of adequate processing models. For instance, Miller and Chomsky's (1963) initial approach of modeling language comprehension as the grammar-equivalent automaton cannot guarantee incrementality, and they were aware of that. The push-down automaton for a context-free language, for instance, would time and again stack up its push down store, thus delaying structural decisions.

Various solutions have been proposed for grammars to handle incrementality in language use. The very first one (to my knowledge) was the Incremental

Procedural Grammar by Kempen and Hoenkamp (1987). This was still a string grammar, but designed to account for the speaker's incremental sentence generation. I used it in *Speaking* (1989). But then, tree grammars took over. Kempen & Harbusch's (1998) Performance Grammar (PG) was, again, explicitly constructed for handling incrementality, in both speaker and listener models. Incrementality is naturally implemented in the linearization component of the grammar, which is essentially a finite state device operating on trees. Meanwhile, various applications of PG have seen the light. One recent example is the modeling of the speaker's generation of clausal coordination and coordinate ellipsis, with all of its gapping and reduction complexities (Kempen, 2009). Another one (Vosse & Kempen 2008) is the implementation of PG in an incremental, but parallel parser (called SINUS). It is parallel in that it can simultaneously entertain different unification alternatives for the same lexical input. At any moment these alternatives are in different states of activation and activations are continuously adapted as new input arrives. The final parse of a sentence corresponds to the configuration of 'winning' (most highly activated) unifications at the end of the sentence. There is no backtracking in the sense of retracing to an earlier point in the sentence and from that point onward selecting an analysis/interpretation that was not entertained before. The claim is that states of activation of unifications are reflected in on-line measures of comprehension load, such as ERP and eye tracking data.

As already noticed above, there exist incremental parsers for TAG. Ferreira (2000) introduced the TAG architecture in her model of the speaker's syntactic production. Ferreira et al. (2004) used it in their account of listener's processing of disfluences in speech. Joshi (1985) himself used aspects of TAG for modeling incremental code switching between Marathi and English. Webber et al. (2003) used TAG in their study of anaphora. See Joshi (2004) for an overview of TAG applications.

Hale (2001) modeled incremental, eager' parsing by way of a probabilistic context-free grammar (based on a sample of the Treebank Corpus), implemented in Stolcke's (1997) probabilistic Early parser. For each next word in the sentence this algorithm computes 1 minus the so-called 'prefix-probability', that is the amount of disconfirmation of (probabilistic) expectations that word provides. That is the word's 'surprise value', which can serve as a measure for the effort it takes to eagerly' or fully exploit the information provided by that word. This measure peaks when reaching the word *fell* in *the horse raced past the barn fell*. More generally, it provides detailed predictions for word-by-word reading latencies. Levy (in press) supplies a rich application of his own, equivalent 'surprise' measure to a range of linguistic cases and experimental data.

Learnability

All formal work on the learnability of grammars, grammatical inference, goes back to Gold's (1967) seminal paper. Under Gold's specific definitions, the somewhat shocking finding was that only finite languages are learnable from so-called 'text presentation, which is an enumeration (infinite for infinite languages) of the sentences of the language. Learnability of infinite languages only exists under 'informant presentation, any enumeration of both the grammatical and the ungrammatical strings over the language's vocabulary (and marked for their (un) grammaticality). With that type of presentation, languages in the Chomsky hierarchy up to context-sensitive (and in addition primitive recursive ones) are learnable. These formal results (reviewed in *FL* I, chapter 8) substantially sharpened the reasoning about Chomsky's *Language Acquisition Device* (LAD), the potential mechanism that would enable any child to infer a grammar for its native language from the linguistic (or other) input received. The fact that children usually do acquire their native language argues for the existence of such a device. I thoroughly treated these matters in *FL* III, Chapter 4. The major issues discussed there are as relevant today as they were in the early seventies. Just to mention some of them: If learnability requires informant presentation, how much negative evidence is (in whatever way) presented to children? The dominant view at that time was: none. No child is told: 'the utterance you (or I) just produced is ungrammatical'. And the fast conclusion was: because a natural language is not learnable, it must be largely innate. That is Chomsky's Universal Grammar (UG). Meanwhile, convincing evidence has been obtained for 'negative evidence' provided by adults to children. Chouinard and Clark (2003), for instance, reported evidence of systematic corrections by adults of children's utterances and evidence of children's attending to, acknowledging and incorporating these corrections.

Another issue was and is: noise ruins learnability. If only one sentence doesn't show up in the limit or if, in informant presentation, one ungrammatical string is marked as grammatical, learnability breaks down in Gold's algorithm. The child's language input is, obviously, quite noisy; how to deal with that? Horning (1969) was the first to conceive of a procedure for selecting or 'learning' a probabilistic grammar from stochastic text presentation. As I reviewed in *FL* I 8.4, probabilistic non-ambiguous context-free grammars are learnable this way (under Homing's definition). One advantage of statistical learning algorithms is that they can be noise-resistant. Meanwhile the statistical modeling of language learning has made substantial advances, in particular by the work of Valiant (1984) and Haussler (1996). Neural network modeling has become a major new statistical approach to issues of language inference (cf. Elman (2005). But there, the perennial problem is that learnability is at best demonstrated by computer simulation; it is never *proven* within this paradigm. It doesn't meet Gold's golden standard. For reviews of these and other issues in learnability, see Jain et al. (1999), Pullum (2003) and Scholz and Pullum (2006).

A potentially important type of linguistic input to language learning children is the prosody of the utterance. That prosody reveals to some extend the syntactic structure of the utterance. That would make tree automata interesting devices in the modeling of grammatical inference. Another good reason for using tree automata in modeling children's early speech corpora is the ubiquitous use of constructions. These are complete holophrases (such as *in there)* or phrases with just one or a small number of variable positions (such as *where N go?).* As mentioned above, tree grammars can naturally handle such fixed expressions by way of elementary trees with multiple lexical anchors. Borensztaijn et al. (2008) have applied this to the Adam, Eve and Sarah corpora from Brown (1973). The developmental notion here is that lexical anchors in the initial multiple-anchored trees are replaced by variables in the course of development. The fixed construction becomes less and less 'fixed', slowly approximating the adult lexical trees, which usually have a single lexical anchor plus a number of variable positions marked for substitution.

Conclusion

When I wrote *Formal Grammars,* the world of formal paradigms in linguistics and its applications was still surveyable, if not simple. That paradisaical situation is long gone. No single linguist or psycholinguist can now oversee the richness of formal devices used in the theoretical and empirical study of natural language and its uses. Using them has become team work. And, at least in psycholinguistics, the use of formal devices has become eclectic. Although trends still come and go, no single approach achieves the aura of being 'the right one'. That is, by and large, a healthy situation. Still, the drive to do things formally right will always be with us, students of language and its uses.

References

Ajdukiewicz (1935). Die syntaktische Konnexität. *Stud. Philos., 1,*1-27.

Borensztajn, G., Zuidema, W., & Bod, R. (2008). Children's grammars grow more abstract with age - Evidence from an automatic procedure for identifying the productive units of language. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci2008),* Washington DC, July 2008.

Bresnan, J. (2006). Is knowledge of syntax probabilistic? Experiments with the English dative alternation. Proceedings of the International Conference on Linguistic Evidence, Tuebingen, 2-4 February 2006, *Roots: Linguistics in search of its evidential base*. Series: Studies in Generative Grammar, edited by Sam Featherston and Wolfgang Sternefeld. Berlin: Mouton de Gruyter.

Büchi (1960). Weak second-order arithmetic and finite automata. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik,* 6,66-92.

Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: MIT Press.

Chomsky, N. (1982). *The generative enterprise. A discussion with Riny Huybregts and Henk van Riemsdijk*. Dordrecht: Foris Publications.

Chouinard, M. and Clark, E. (2004). Adult reformulations of child errors as negative evidence. *Journal of Child Language,* 30, 637-669.

Comon, H., Dauchet, M., Gilleron, R., Jacquemard, F, Lugiez, D., Löding, Ch., Tison, S., & Tommasi.M. (2007) *Tree Automata Techniques and Applications*. E-Bookhttp://tata.gforge.inria.fr/

Gécseg, F. & Steinby, M. (1997). Tree languages. In: G. Rozenberg & A. Salomaa (Edss.). *Handbook of formal languages*. Vol. 3 *Beyond words*, 1-68.

Gold, E.M. (1967). Language identification in the limit. *Information and Control, 10,*441-474.

Ellis, C.A. (1971). Probabilistic tree automata. *Information and Control, 19,* 401-416.

Elman, J. L. (2005). Connectionist models of cognitive development: Where next? *Trends in Cognitive Sciences,* 9,111 -117.

Ferreira, F. (2000). Syntax in language production: An approach using Tree-Adjoining Grammars. In L. Wheeldon (Ed.), *Aspects of language production*. Cambridge, MA: MIT Press.

Ferreira, F, Lau, E.F. & Bailey, K.G.D. (2004). Disfluencies, language comprehension, and Tree Adjoining Grammars. *Cognitive Science, 28,* 721-749.

Fodor, J.A., Garrett, M.R, Walker, E.C.T. & Parkes, CH. Against definitions. *Cognition, 8,* 263-367.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL, 2,* 159-166.

Harbusch. K. & Kempen, G. (2002). A quantitative model of word order and movement in English, Dutch and German complement constructions. In: *Proceedings of the 19th International Conference on Computational Linguistics (COLING-2002),* Taipei (Taiwan). San Francisco: Morgan Kaufmann.

Haussler, D. (1996). Probably aproximately correct learning and decision-theoretic generalizations. In: P. Smolensky, M.C. Mozer, & D.E. Rumelhart. (Eds.), *Mathematical perspectives on neural networks,* 651-718. Mahwah, NJ: Erlbaum.

Horning, J.J. (1969). A study of grammatical inference. *Technical Report CS 139. Stanford Artificial Intelligence Project).*

Jackendoff, R. (2002). *Foundations of language*. Cambridge, MA: MIT Press.

Jain, S., Osherson, D.N., Royer, J.S., & Sharma, A. (1999). *Systems that learn: An introduction to learning theory*. Second edition. Cambridge, MA: MIT Press.

Joshi, A. (1985). Processing of sentences with intra-sentential code-switching. J.Horeck (Ed.), *Proceedings of COLING 82*. Amsterdam: North-Holland Publishing Company.

Joshi, A. K. (2004). Starting with complex primitives pays off: complicate locally, simplify globally. *Cognitive Science, 28,* 637-668.

Joshi, A.K. & Shabes, Y. (1997). Tree-adjoining grammars. In: G. Rozenberg & A. Salomaa (Eds.). *Handbook of formal languages.* Vol. 3 *Beyond words,* 69-123.

Kaplan, R.M. & Bresnan, J. (1982). Lexical-functional grammar: A formal system for grammatical representation. In: J. Bresnan Ed.), *The mental representation of grammatical relations,* 173-281. Cambridge MA: MIT Press.

Kempen, G. & Harbusch, K. (1998). Tree Adjoining Grammar without Adjoining? The case of scrambling in German. *Proceedings of TAG+4 - Fourth International Workshop on Tree Adjoining Grammars and Related Formalisms.* Philadelphia, Pennsylvania/USA: IRCS Report #98-12,80-83.

Kempen, G., & Harbusch, K. (2002). Performance Grammar: A declarative definition. In: A. Nijholt, M. Theune & H. Hondorp (Eds.), *Computational Linguistics in The Netherlands 2001.* Amsterdam: Rodopi.

Kempen, G. & Harbusch, K. (2005). The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In: S. Kepser & M. Reis (Eds.), *Linguistic Evidence—Empirical, Theoretical, and Computational Perspectives.* Berlin: Mouton De Gruyter.

Kempen, G. & Harbusch, K. (2008). Comparing linguistic judgments and corpus frequencies as windows on grammatical competence: A study of argument linearization in German clauses. In: Steube, Anita (Ed.), *The discourse potential of underspecified structures,* 179-192. Berlin: DeGruyter.

Kempen, G. & Hoenkamp (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science, 11,*201-258.

Kempen, G. (in press 2009). Clausal coordination and coordinate ellipsis in a model of the speaker. *Linguistics.*

Klein, W. (1974). *Variation in der Sprache. Ein Verfahren zu einer Beschreibung.* Kronberg: Scriptor Verlag.

Knight, K. & Graehl, J. (2005). An overview of probabilistic tree transducers for natural language processing. In: *Proceedings of the Sixth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing).* Lecture Notes in Computer Science. Springer Verlag, 2005.

Levelt, W.J.M. (1989). *Speaking. From intention to articulation.* Cambridge, MA: MIT Press.

Levinson, S.C. (2000). *Presumptive meanings.* Cambridge, MA: MIT Press.

Levy, R. (in press). Expectation-based syntactic comprehension. *Cognition.*

Miller, G.A. & Chomsky, N. (1963). Finitary models of language users. In: R.D. Luce, R.R. Bush & E. Galanter, *Handbook of mathematical psychology.* New York: John Wiley.

Montague, R. (1970). English as a formal language. Reprinted in: R. Thomassen (Ed.), (1974), *Formal philosophy: selected papers of Richard Montague,* 188-221. New Haven: Yale University Press.

Partee, B.H., with H. Hendriks (1997). Montague grammar. In: J. van Benthem & A. ter Meulen (Eds.). *Handbook of logic & language.* 5-91. Amsterdam: Elsevier and Cambridge, MA: MIT Press.

Peters, P.S. & Ritchie, R.W. (1973). On the generative power of transformational grammars. *Information Sciences, 6,*49-83.

Pullum, Geoffrey K. (2003) 'Learnability.' *The Oxford International Encyclopedia of Linguistics.* Second edition, 431-434. Oxford: Oxford University Press.

Resnick (1992). Probabilistic tree-adjoining grammars as framework for statistical natural language processing. In: *Proceedings of the 15th International Conference on Computational Linguistics (Coling'92).* Nantes.

Schabes, Y. (1992). Stochastic tree-adjoining grammars. In: *Proceedings of the 15th International Conference on Computational Linguistics (Coling'92).* Nantes.

Scholz, B.C. and Pullum, G.K. (2006) Irrational nativist exuberance. In: R. Stainton (ed.), *Contemporary Debates in Cognitive Science,* 59-80. Oxford: Basil Blackwell.

Seuren, P.A.M. (1969). *Operators and nucleus. A contribution to the theory of grammar.* Cambridge: University Press.

Seuren, P.A.M. (2009). *Language from within. 2* Vols. Oxford: Blackwell Publishers.

Sprenger, S.A., Levelt, W.J.M., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language,* 54,161-184.

Stabler, E.P. (2004). Varieties of crossed dependencies: structure dependencies and mild context-sensitivity. *Cognitive Science, 28,* 699-720.

Steedman, M. (2000). *The Syntactic Process.* Cambridge, MA: MIT Press.

Stolcke, A. (1979). Linguistic knowledge and empirical methods in speech generation. *AJ Magazine,* 18,25-31.

Suppes, P. (1970). Probabilistic grammars for natural language. *Synthese, 22,* 95-116.

Valiant, L. (1984). A theory of the learnable. *Communications of the ACM, 27,*1134-1142.

Vosse, Th. & Kempen, G. (2008). Parsing verb-final clauses in German: Garden-path and ERP effects modeled by a parallel dynamic parser. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci2008),* Washington DC, July 2008.

Valiant, L. (1984). A theory of the learnable. *Communications of the ACM, 27,*1134-1142.

Webber, B., Stone, M., Joshi, A.K. & Knott, A. (2003). Anaphora and discourse structure. *Computational Linguistics, 29,* 545-588.