

Using Stochastic Tree-Substitution Grammar in Iterated Learning Simulations as a way of Approaching Issues in Diachronic Syntax

Dave Cochran, MSc Developmental Linguistics programme, Theoretical and Applied Linguistics,
University of Edinburgh*

Stochastic Tree-Substitution Grammar (STSG) offers a compelling, emergentistic alternative to classical symbolic rule- or constraint-governed grammars, in which a large parsed corpus may be used as an active grammar with a simple procedure for the incremental recombination of tree-fragments extracted from the training corpus, combined with statistical disambiguation to keep check on overgeneration (Bod 1998, 2003). The formalism originates in computational linguistics as an NLP technology, but it has lately been increasingly advanced as a descriptive theory of human syntactic processing (Bod, Hay and Jannedy 2003). Iterated Learning Simulations (ILS) are a class of computational simulation whereby successive generations of virtual agents perform tasks, (in the linguistic case generally using a greatly simplified toy language), having learnt the behaviour from the performance of previous generations, and are used extensively to model hypotheses regarding the evolution of language (Briscoe 2002, Hurford 2000, Kirby 1999). It is my hope that the much richer, more realistic computational grammars achievable by STSG will allow for the computational modelling of processes of grammatical change in ILS typologically analogous to documented cases of real-world grammatical change, and that eventually such simulations may furnish theoretical tools for the explanation of real-world grammatical change. This paper details some preliminary ideas for this research project, and some of the major challenges involved.

1. Introduction

The following paper details my plans for a large research project, too large to be wholly covered in my MSc thesis, or even my PhD, and which has occupied my thought since 2002, when I read *Beyond Grammar: An Experience-Based Theory of Language*, by Rens Bod (Bod 1998), in which the Stochastic Tree-Substitution Grammar (henceforward STSG) formalism is described. STSG offers a compelling, emergentistic alternative to classical symbolic, rule-governed grammars, in which a large parsed corpus (or *treebank*) may be exploited as an active grammar, using a simple procedure for the incremental recombination of tree-fragments extracted from the training corpus and statistical disambiguation to keep check on overgeneration. The formalism originates in Natural Language Processing research, and as such the considerable majority of the STSG research has been concerned with its use for parsing – Data Oriented Parsing or DOP. For a more detailed technical exposition, see below

Iterated Learning Simulations (ILS) are a class of computational simulation whereby successive generations of virtual agents perform tasks, (in the linguistic case generally using a greatly simplified toy language), having learnt the behaviour from the performance of previous generations, and continuing until their behaviour settles into a

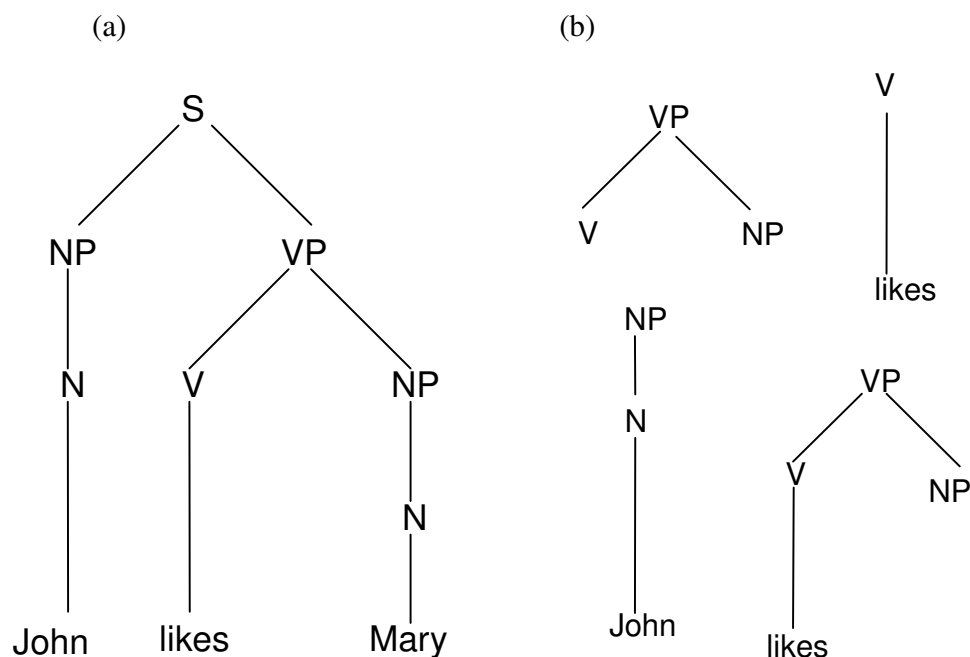
* This paper was presented to the MSc Developmental Linguistics Seminar, Theoretical and Applied Linguistics, University of Edinburgh, on 27th May 2005, under the title “Can a rule-free grammar explain diachronic rule-change? A proposal for an application of Stochastic Tree-Substitution Grammar in Iterated Learning Simulations” and in abridged form to the College of Arts and Social Sciences Postgraduate Conference at the University of Aberdeen on 23rd June 2005.

stable attractor, or an arbitrary number of (thousands of) generations has been reached. It is used extensively to model hypotheses regarding the evolution of language (Briscoe 2002, Hurford 2000, Kirby 1999), but it is my hope that the much richer, more realistic computational grammars achievable by STSG will allow ILS's to model processes of grammatical change typologically analogous to common types of documented real-world grammatical change, such as grammaticalisation. My long-term goal is to develop computational, statistical & theoretical tools which may be brought to real historical data and used to explain real cases of grammatical change such as the vagaries of the English determiners or the English modals in the Middle Ages.

2. DOP1

The simplest form of DOP is DOP1, as described in Bod 1998 (p12-23 and 40-50). The parser uses a large parsed corpus¹ divided into a training corpus and a much smaller corpus against which the parser is tested. The parser takes every tree in the training corpus and breaks it down into all of its possible subtrees, according to certain wellformedness rules (see 1-3 below and figure 1c).

- 1 Every subtree must have at least one link
- 2 Every link must have a node on either end
- 3 Sister relationships must be preserved



¹ Such as the the Penn Treebanks (in English, Chinese, Arabic, etc) or the "Developing a Morphologically and Syntactically Annotated Treebank Corpus For Turkish" Project sponsored by the METU Informatics Institute & Sabanci University

(c)



Figure 1: (a) the parse tree for “John likes Mary” (b) well-formed subtrees of “John likes Mary” (c) ill-formed subtrees of “John likes Mary”.

The parser is given strings from the test corpus and builds up new, hypothetical parse-trees for these strings using the fragments available to it from the training corpus, starting with a fragment with an S-node at the top, and then, for each nonterminal leaf-node, working rightwards, substitutes in additional subtrees, the topmost node of which must carry the same label as the node to be substituted. (see figure 2).

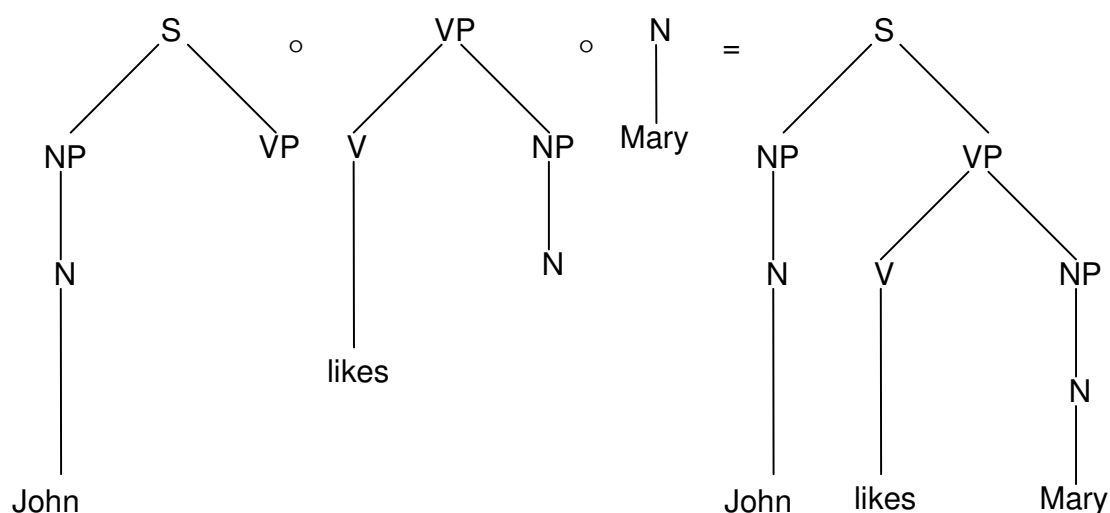


Figure 2: A derivation of “John likes Mary”. \circ is the operator for the tree-substitution operation.

For each possible parse, there will be several possible derivations, and for some sentences, multiple parses may be possible, depending on the contents of the training corpus. For each subtree used in a derivation, its probability is calculated as its total frequency of occurrence over the number of subtrees with the same root node

$$P(t) = \frac{|t|}{\sum_{t':r(t')=r(t)} |t'|} \quad (\text{eqn 1})^2$$

² t = subtree

The probability of a derivation is the product of the probabilities of its subtrees, and

$$P(t_1 \circ \dots \circ t_n) = \prod_i P(t_i) \quad (\text{eqn 2})$$

The probability of a parse is the sum of the probabilities of its possible derivations

$$P(T) = \sum_{D \text{ derives } T} P(D) \quad (\text{eqn 3})^3$$

The probability of a parse in relation to the input string is the probability of the parse over the summed probabilities of all parses yielding that same input string. Where unambiguous, $P=1$

$$P(T \mid T \text{ yields } W) = \frac{P(T)}{\sum_{T' \text{ yields } W} P(T')} \quad (\text{eqn 4})^4$$

The output of the parser is most probable parse. Obviously, calculating every parse and derivation would be computationally prohibitive, so shortcuts are taken. Monte Carlo disambiguation takes n many random samples the set of possible derivations and treats them as being representative of the whole set, with the likelihood of sampling error decreasing with any increase of the value of n . Bod (ibid p.54) reports accuracies of 85% on the ATIS⁵ corpus.

Other versions of DOP do even better, such as Simplicity-Likelihood DOP (SL-DOP), as developed and tested for music and language in Bod 2003. Rather than simply picking the most likely parse, Monte-Carlo disambiguation in SL-DOP first picked the n most likely parses, and then of these, outputted the simplest (the simplest parse being that with the shortest derivation). It was tested for language on Penn Wall Street Journal Corpus and for music on Essen Folk Song Collection, and the optimum value of n was found to be 12. It was assessed using PARSEVAL Mean F-scores⁶, and at $n = 12$;

- For WSJ, F = 90.2%
- For Essen, F = 87.3%

What is particularly important about these results is that they show that Stochastic Tree Substitution (STS) has other applications than language. In the same paper Bod laments the lack of a suitable treebank of pictures upon which to try out SL-DOP for the

r = root node of subtree

³ T = tree

D = derivation

⁴ W = word string

⁵ Air Transport Information System – part of the Penn Treebank.

⁶ The F-score is a measure balanced between precision – the percentage of constituents of the computer's parse which are correct, and recall – the percentage of the constituents of the correct parse that the computer got)

parsing of images, and elsewhere (Bod 2004) uses a similar formalism to describe processes of scientific reasoning. At the other end of the scale, I would intuit that motor-memory could also be usefully analysed with STS. From here the question follows, how much of cognition can STS model?

However, two major technical obstacles must be overcome to allow the application of STSG in ILS. Systems for sentence generation and bootstrapping⁷, to the best of my knowledge, have yet to be developed. Note that there is a need for realism here – the goal is not the chasing of theoretical generalities, but rather a contribution to historical linguistics. If the goal were to see what, minimally, the existence of certain diachronic typologies requires us to hold of how humans cognise and process language, and of the dynamics of human communication, a maximally simplified and abstract model would not only be acceptable but desirable. But instead the ultimate goal is to explain, in detail, the documented development of real-world language change, on the basis of our best knowledge of the cognition and processing of language and the dynamics of human communication, and this requires models that incorporate as much of that knowledge as possible.

3. Data-Oriented Generation

It is my intention to develop a rudimentary Data-Oriented Generation (DOG) system as the basis of my MSc thesis. This is an ambitious project, not only because it hasn't been done before, but also because my goal in so doing is not simply to devise and implement an algorithm that works, but to develop an STSG-based theory of human sentence composition, and as such empirical research will be necessary in order to ensure psycholinguistic validity. One particular empirical question I wish to tackle is that of the input to sentence production; as such, I propose to develop two models – a “mediated” model (“MDOG”), in which experience of the world is encoded in a predicate logic-like “language of thought”, and sentence generation “translates” these expressions into natural language; and an “unmediated” model (“UDOG”), in which the mapping between experience and language is more direct.

Typically, while psycholinguists studying language production have assumed the existence of pre-linguistic messages providing the input for language production, philosophers have had grave misgivings about this. Typical of this latter trend is Simon Blackburn's argument (1984, p40-67) that the Pre-Linguistic Message Hypothesis (or “Language of Thought Hypothesis” as it is called in the philosophical literature) is a *dog-legged theory*, insofar as our ability to understand and perform operations with pre-linguistic messages stands in want of explanation in precisely the same way as our ability to understand expressions in language, thus inviting an infinite regress. The argument at the heart of all this is Wittgenstein's example of understanding the meaning of the word “red”, in which it is asserted that, if one supposes the mind to contain an image of “red” to provide the word “red” with meaning, it is functionally equivalent to having an image

⁷ i.e. the first extraction of structure and meaning from an unsegmented speech stream by a learner lacking any previous linguistic knowledge.

of “red” *outside* the mind to correlate word to meaning, say, a labelled card with a red patch painted on it.

When I hear the order "fetch me [a red flower from that meadow]." I draw my finger across the chart from the word "red" to a certain square, and I go and look for a flower which has the same colour as the square. ... [But] consider the order "imagine a red patch". You are not tempted in this case to think that before obeying you must have imagined a red patch to serve you as a pattern for the red patch which you were ordered to imagine.

Wittgenstein 1969, p3.

In contrast, psycholinguists have, without much discussion, tended to favour the assumption of pre-linguistic messages as a way of abstracting their desired object of investigation, the complex of systems by which we select the words, inflections, syntactic structures, phonemes and suchlike with which our desired meaning is to be expressed away from a matter which is murkier, more daunting, and less accessible to experimental research, that of how, in the first place, we decide which meanings we want to express. Levelt (1989), for instance, divides his “blueprint for the speaker” (p.9) first of all into the “Conceptualizer” and the “Formulator”. However, it is quickly apparent that while the Formulator is a fairly clearly defined set of linked subsystems for handling different layers of the surface structure of linguistic expressions, the Conceptualizer, by Levelt’s own admission, is a sort of heterogeneous “not-the-liver” category⁸ set up to do everything the Formulator doesn’t.

The sum total of these mental activities will be called *conceptualising*, and the subserving processing system will be called the *Conceptualizer* (in full awareness that this is a reification in need of further explanation – we are, of course, dealing with a highly open-ended system involving quite heterogeneous aspects of the speaker as an acting person)

p.9, author’s emphasis.

We find in Levelt’s further exposition of the Conceptualizer (p70-106) that the output of “messages” which it feeds into the Formulator must meet certain criteria - they must be “propositional” (in a broad sense) (p.72-96), they must have perspective (i.e. carry information about topicality, news value, etc) (p96-100) and mood (eg. interrogative, declarative, etc) (p100-103), and be marked for whatever supplementary information the grammar of the language in question demands (eg. information about evidentiality is optional in English, but mandatory in Karaja (Maia, 2000)). Levelt laments the absence of a “message grammar”, or any immediate prospect of one (p70). At this point, the “message” seems so much like language itself, that one or the other must surely be redundant.

However, the point of the philosophical argument here is not that logic-like pre-linguistic messages *cannot* exist, just that we cannot simply infer their existence from the gross facts of our language use and cognition. But a “language of thought” certainly wouldn’t

⁸ I am indebted to Bedford (1997) for this singularly useful expression. It refers to a particular kind of fallacious category in cognitive science – one by which the discovery of a genuine category or “organ” within cognition is supplemented by the putative discovery of a second category comprising everything that the first doesn’t.

be redundant in a pre-linguistic animal, for which a system of categorial scene-representations would facilitate the anticipation of forthcoming events, and evolution has no problem with the retention of redundant systems, at least if they have no negative impact on the survival of an animal - indeed, they may add robustness. Furthermore, there is evidence that just such a system existed in our prelinguistic ancestors, and continues in modern nonhuman primates: Hurford (2003) cites a wealth of neurological evidence to the effect that the processing of vision (and, in thumbnail sketch, other senses) may be divided into two main channels – dorsal and ventral, which may crudely be characterized as having “where” and “what” functionalities. That is to say, the dorsal channel locates entities in egocentric space, while the ventral tells you what they are and what they’re doing. Hurford hypothesises that the entities whose position the dorsal channel tracks correspond to arguments (variables, x , y , etc) in predicate logic, and the ventral channel applies (first-order) predicates to them.

With this in mind, I now wish to outline MDOG and UDOG.⁹ Both models are intended to produce declarative sentences and questions regarding visual stimuli. In both cases the picture inputs will be analysed with a data oriented parser. In laying out nested spatial relations within the picture, these tree-structures amount to the output of a pre-conceptual “where” task. In other words, the DOPP does precisely the job of the dorsal channel, and I hypothesise that this is exactly what the dorsal channel is, at least with regard to vision – a Data-Oriented Picture Parser. A fully fleshed, realistic DOPP would be a somewhat different kind of creature to DOP for string inputs (i.e. language or music). In the latter case the material to be parsed is arranged on a single dimension – time - and so the only primitive difference-making relation on that dimension is “before/after”, whereas in the former, even restricting the parser merely to 2D arrangements would require a potentially infinite set of relations, in order that, for example, the subtrees in figure 3 should all be considered as distinct from one another in the calculation of probabilities. In practice, types of spatial relationship in picture parses may need to be grouped within a set of fuzzy categories. However, such a system may become baroque in its complexity, and it may be necessary to, temporarily at least, restrict the system to 1D visual inputs, and to parsing groupings of discrete objects rather than features within objects.

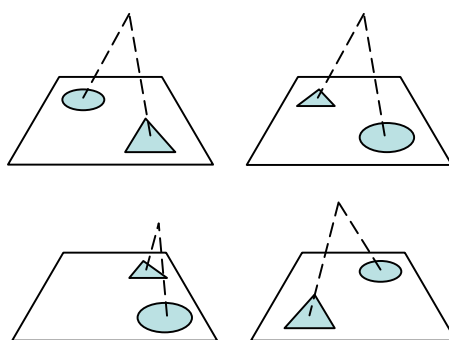


Figure 3: Distinct subtrees in Data-Oriented Picture Parsing.

⁹ Please note that if any of the technical details at this point are hazy, it may well be because I’m glossing over something I haven’t quite worked out for myself – things that will probably need an element of trial and error to get right.

MDOG and UDOG differ according to the way the picture and word-string inputs are related to one another: MDOG takes advantage of STSG's ability to handle trees decorated with predicate logic expressions. To give two instances of how this may be done¹⁰, in the toy corpus illustrated in figure 4.a, expressions like $\exists x(\text{MAN}x \ \& \ \text{WHISTLES}x)$ are located at the root node, and broken down with lambda-abstractions as you work down to the terminal nodes (van den Berg, Bod and Scha 1994, cited in Bod Bonemma and Scha 1996), whereas in figure 4.b, the expressions above the immediate parents of the terminal nodes are replaced with more abstract substitution-schemas (Bod, Bonnema and Scha 1996).

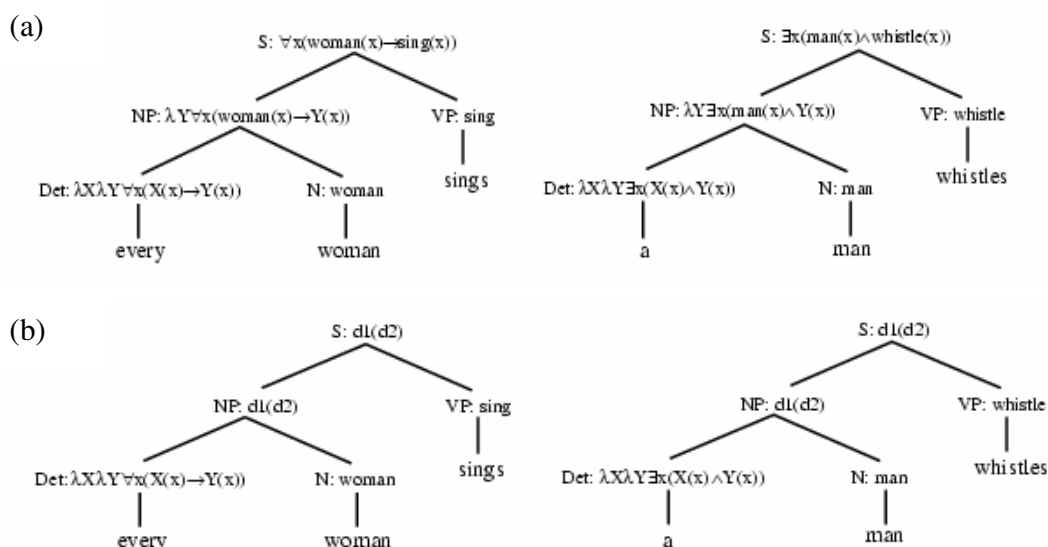


Fig 4: Two toy corpora from; R. Bod, R. Bonnema and R. Scha, 1996. “A Data-Oriented Approach to Semantic Interpretation.” *Proceedings Workshop on Corpus-Oriented Semantic Analysis, ECAI-96*, Budapest, Hungary. Bod et al note that constraining DOP to process semantic annotations of the type shown in figure 4(b) actually improves the parser's accuracy for syntax, and its overall processing speed.

My plan for MDOG is to decorate the trees in two separate corpora, of images and of sentences describing images, with First Order Predicate Logic (FOPL) expressions – the input to the MDOG will be the FOPL expression output of the DOPP. The same composition process as in DOP follows, but instead of being constrained to yield a specified word-string output, it will be constrained to yield the specified FOPL expression output, and overgeneration will be handled by the same Monte-Carlo disambiguation. MDOG embodies the hypothesis that the ventral stream is the system that decorates the DOPP trees, and the language faculty has evolved to take inputs from this “PREDICATE(x)” system.

¹⁰ It is not necessary at present to determine which way is best to decorate the trees used in MDOG.

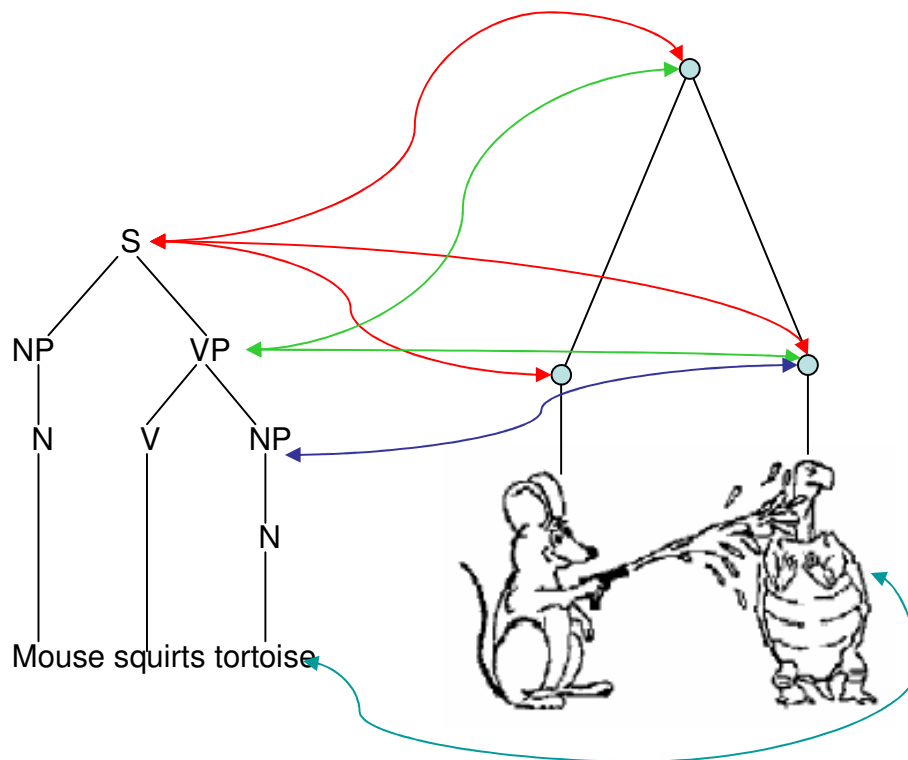


Figure 5: paired parse trees with selected dynamic links

UDOG differs from this in that, rather than operating with two separate corpora, it will use a single training corpus of picture-description pairings¹¹, in which the paired parses are joined at multiple nodes by dynamic links as illustrated in figure 5. The difference lies in the subtree-decomposition process: it will be necessary to evaluate different sets of wellformedness rules by trial-and-error, the common core of the possible rulesets being that if a well-formed subtree of a picture parse contains one or more nodes connected to one or more dynamic links, all dynamic links must be included and must have well formed subtrees of the corresponding sentence on the other end.

The consequence is that a parse of a novel picture will yield a set of language-subtrees. The same composition process as in DOP follows, but instead of being constrained to yield a specified wordstring output, it will be constrained to include the provided subtrees, and overgeneration will be handled by the same Monte-Carlo disambiguation.

The evolutionary corollary of this would be that in humans the ventral channel has in fact been wholly exapted into the language processing system, perhaps in the role of managing the system of dynamic links. This presents our cognitive engagement with the world as a being structured and mediated to a great extent by language, in a way which may seem familiar to advocates of the Sapir-Whorf hypothesis and certain continental philosophers. Prelinguistic infants may experience PREDICATE(x) as our prelinguistic

¹¹ though any given picture will be entered with multiple descriptions, depending on topic, focus, etc, and descriptions will be recycled if they are truthfully applicable to more than one picture

ancestors, due to the lack of exemplars in the linguistic part of their internalised treebank to which to attach dynamic links, but over the process of L1 acquisition this would be crowded out by dynamic link-formation, as the linguistic exemplar-base expands. One might expect the presence of these avatistic non-linguistic predicates to have some shaping or distorting effect on dynamic link formation, and thus natural language semantics.

The last part of my MSc thesis, I'm afraid is something I can't as yet say much about, and that is assessing the realism of MDOG and UDOG. The criterion isn't actually which one works best – I expect them both to work – but which one most resembles human performance – bearing in mind, of course, the abstractions involved in their devising. But what predictions each model will make about human performance (eg, error-making, developmental schedules, etc), and therefore what experimental confirmation may be required, won't be known until they're actually working.

4. Bootstrapping

The other major gap in STSG research that must be filled in order to run Iterated Learning Simulations is the need for a bootstrapping process. I anticipate this to be the first major challenge of my PhD, and the question is likely to be affected by which DOG I adopt.¹² The various DOPs in circulation and both my DOGs all rely on having a set of exemplars already parsed and annotated to serve as a training corpus, but in the case of the human infant, where does that come from? The question of bootstrapping may be divided into two subquestions;

1. How does the infant segment and group the speech-stream into tree-structures?
2. How does the infant apply linguistic categories to the various nodes of these structures?

I only have time for some very loose suggestions as to how to tackle these questions, but I would hazard that when an infant first starts to recognise frequently heard segments, the recognition of which is perhaps repeatedly reinforced by the presence of a physical object (such as a ball), the recognised segment plus its context is stored as a tree structure similar to figure 6

¹² having some sort of DOG is a prerequisite for the development of a realistic STSG-bootstrapper, given the importance of children's babbling and other linguistic play in their language development. For one thing, each time a DOLP produces or processes another utterance, that utterance is then added to the training corpus. A child expands her linguistic resources and renders them more robust (for instance against parental correction) simply by playing around with the resources she already has.

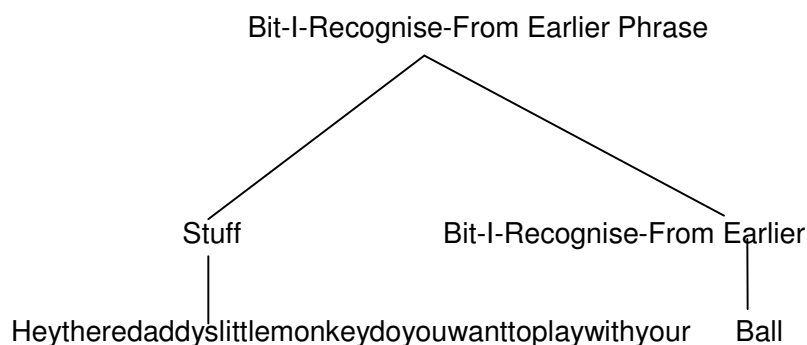


Figure 6: Baby's first tree structure - note the head-complement relationship between the recognized element and the unfamiliar material.

The latter question, regarding categoriality, is more of a vexed issue, especially since it is an open question whether part-of-speech tagging is actually necessary. The parsing of music in Bod 2002 achieved PARSEVAL scores of up to 87.3% without any node-labelling. Barner and Bale 2002 present psycholinguistic arguments for incorporating morphological word-formation into the syntax, without any qualitative difference from the formation of phrases – thus a nonspecified root becomes a noun or verb when inserted into a nominal or verbal context. My intuition is that the logic-like annotations of MDOG, or the dynamic links of UDOG, should in fact prove sufficient to do the job part-of-speech tags do in DOP1. This incidentally is borne out by an evolutionary simulation by John Batali (Batali 2002) using a formalism similar to STSG. That dynamic links can perform predicate-like role assignments should be clear from figure 5: the number of dynamic links leading from each sentence tree-node, and the level within the picture tree to which they connect, correlate closely to the role played in the sentence by the phrase or word under the node in the sentence tree.

5. STSG and the lexicon

One interesting corollary of all this is that, whatever the relationship between language and the ventral stream, if the attachment of semantic information¹³ to the word-string treebank is sufficiently rich, and efficiently exploited by the processor, lexicon, like grammar, becomes wholly obsolete. It may further be noted that such “quasi-lexical” semantic information may also be attached to higher-level nodes, allowing the system to deal consistently with both prototypical “lexical items” and “lexicalized” constructions, thus making it a form of Construction Grammar.

6. Dynamic systems theory

The last question I want to deal with is how I actually hope to extract theoretical, statistical, and computational tools for the analysis of real diachronic change out of this. Because what's clear is that if STSG is, as I believe it to be, the correct account for the

¹³ Which need only be semantic information specific to the particular utterance in context, rather than an exhaustive account of the possible meanings of a given word or phrase.

human language faculty, then grammatical rules are not only emergent, but epiphenomenal. I define a phenomenon as emergent if and only if the language used to describe it cannot be the same as the language used to describe the underlying causal processes that implement it,¹⁴ and an epiphenomenon as an emergent phenomenon for which the language used to describe the observed *apparent* regularities stands in a “fractal” or “asymptotic” relationship to the underlying causal processes, by which I mean that attempts to describe what’s happening in such terms will approach empirical adequacy asymptotically, only as brittleness, complication and exceptionality tend towards infinity. To quote Bod;

...the knowledge of a speaker/hearer cannot be understood as a grammar, but as a statistical ensemble of language experiences that changes slightly every time a new utterance is perceived or produced. The regularities we observe in language may be viewed as emergent phenomena, but they cannot be summarized into a consistent non-redundant system that unequivocally defines the structure of new utterances ... Language change could be explained as a side-effect of updating the statistical ensemble of language experiences”

Bod 1998, p.145

With this in mind, it is no surprise that generativists prefer to primarily to describe “ideal speaker-hearers in completely homogeneous speech communities” and abstract away from the messiness and approximation of real language performance.

However, it is desirable to have some way of talking about these emergent regularities, and my hope is that this can be achieved by conceiving of the linguistic “statistical ensemble” as a dynamic system, in which more or less stable sets of point attractors exist. One type of attractor that is of particular interest to the diachronic linguist is the catastrophe curve.

¹⁴ I am indebted to Simon Kirby for this definition.

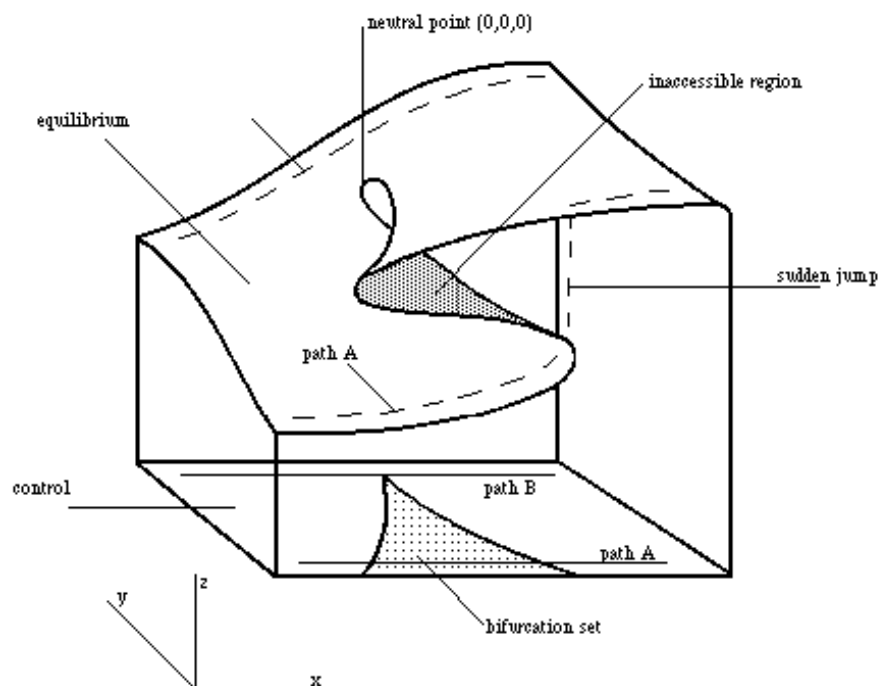


Figure 7: A Cusp Catastrophe: graph taken from Han van Der Maas's webpage: users.fmg.uva.nl/hvandermaas/cata.html. The cusp is a class of catastrophe with two control variables and one behavioural variable.

A catastrophe curve (see figure 7) is the set of equilibrium states of a set of behavioural parameters graphed against a set of control parameters, the “equilibrium surface”, in which areas of bimodality exist, and sudden discontinuities in the value of the behavioural parameter(s) may occur when the values of the control parameters cross out of the bimodality zone, as illustrated in “path A” above. This could have applications for the explanation of discontinuous phenomena in historical syntax¹⁵.

Taking assumptions 1-6 below as a rough sketch, based on my readings of STSG and Catastrophe Theory, of an assumption-set from which such an application may be derived, I propose to test hypotheses 7-9 as the main part of my PhD.

1. In any STSG, the total number of possible trees (P) generable by the set of subtrees (S) of the parse forest far exceeds the number liable to be judged grammatical by the DOLP (G).
2. If $P^1 \dots P^x$ is a set of selected statistical characterisations of members of P , and
3. $S^1 \dots S^y$ is a set of selected statistical characterisations of members of S , then
4. $P^1 \dots P^x$ may be analysed as a topographic space in which every point corresponds to a member of P , and

¹⁵ The vagaries of the English modals, for instance, being the classic testing-ground for the diachronic application of theories hitherto focused on matters synchronic

5. For fixed values of $S^l \dots S^y$, G may be analysed as the set of stable equilibria of $P^l \dots P^x$.
6. If one factors in variation in $S^l \dots S^y$, one requires a topographic space of $x+y$ many dimensions; the $x+y$ -dimensional extension of G is Γ . In this case $S^l \dots S^y$ may be analysed as the control variables, and $P^l \dots P^x$ the behaviour variables.

I hypothesise that;

7. Areas of multimodality in G correspond to free variation within the string-language, and may be analysed on the basis of some form of Stochastic CT (Hartelman 1997).
8. The catastrophe curves bounding areas of multimodality in Γ , considered in terms of Deterministic CT, describe the potential for discontinuous changes in syntax as a consequence of continuous variation in the values of $S^l \dots S^y$, and a specified course of such variation over time may be taken as a particular hypothetical instance of syntactic change.
9. By comparing hypothetical models/instances generated according to the above procedures with statistical characterisations and/or STSG simulations performed on historical corpora, one may, for instance by varying the number and type of statistical characterisations included in $P^l \dots P^x$ and $S^l \dots S^y$, arrive at a “best fit” model for a particular instance of documented historical change.

Please note that in the topographic spaces proposed, there is no dimension of time; this is important, because the variant values of $S^l \dots S^y$ might not occur in different time-states of a single DOLP, but rather in successive generations of DOLPs, each providing the next 1-3 generations with their initial training corpora by the results of their speech-generation routines before “dying”.

7. Further research

Factoring in contact phenomena, sociolinguistic phenomena, and placing the agents within a virtual geographic space that constrains their interaction would form the basis of the next stage of the research, which could, for instance, be directed at illuminating the processes of inflectional degradation in the Danelaw, the area of Northern England subject to Danish Law, and greatly settled by Danes, in the late 9th and early 10th centuries. Real geographical and archaeological data could be incorporated into the model. Ultimately, what I hope to get out of this is an entirely new way of doing historical linguistics.

References

- Barner, D. and A. Bale (2002), “No Nouns, no verbs: psycholinguistic arguments in favor of lexical underspecification” *Lingua* 112, p. 771-791
- Batali, J. (2002), “The Negotiation And Acquisition Of Recursive Grammars As A Result Of Competition Among Exemplars” In Ted Briscoe, editor, *Linguistic Evolution*

- through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.
- Bedford, F (1997). False categories in cognition: The Not-the-Liver fallacy. *Cognition*, 64, 231-248.
 - van den Berg, M., R. Bod and R. Scha, (1994). "A Corpus-Based Approach to Semantic Interpretation", *Proceedings Ninth Amsterdam Colloquium*, Amsterdam, The Netherlands.
 - Blackburn, S (1984). *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Oxford University Press.
 - Bod, R. (1998), *Beyond Grammar; An Experience-Based Theory of Language*, Stanford, California: Centre for the Study of Language and Information.
 - Bod, R (2002) "A Unified Model of Structural Organization in Language and Music" *Journal of Artificial Intelligence Research*, 17, 289-308.
 - Bod, R. (2003), "An Efficient Implementation of a New DOP Model" *Proceedings EACL'03*, Budapest, Hungary.
 - Bod, R. (2004). "Exemplar-Based Explanation." *Proceedings Computation and Philosophy (ECAP'04)*, Pavia, Italy.
 - Bod, R., R. Bonnema and R. Scha, (1996). "A Data-Oriented Approach to Semantic Interpretation." *Proceedings Workshop on Corpus-Oriented Semantic Analysis*, ECAI-96, Budapest, Hungary.
 - Bod, R., J. Hay and S. Jannedy (eds.), 2003. *Probabilistic Linguistics*. Cambridge, Ma: MIT Press.
 - Briscoe, E. J. (2002) "Grammatical Acquisition and Linguistic Selection". In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.
 - Hartelman, P. (1997) *Stochastic Catastrophe Theory*, unpublished PhD thesis, Universiteit van Amsterdam.
 - Hurford, J. (2000) "Social transmission favours linguistic generalization". In Chris Knight, James R. Hurford and Michael Studdert-Kennedy, editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 324-52. Cambridge: Cambridge University Press.
 - Hurford, J (2003), "The Neural Basis of Predicate-Argument Structure" *Behavioural and Brain Sciences*, 26.3, 261-316.
 - Kirby, S. (1999) "Learning, bottlenecks, and infinity: a working model of the evolution of syntactic communication". In K. Dautenhahn and C. L. Nehaniv (eds.), *Proceedings of the AISB'99 Symposium on Imitation in Animals and Artifacts*, pages 55.63. Society of the Study of Artificial Intelligence and the Simulation of Behaviour, 1999.
 - Levelt, W (1989). *Speaking: From Intention to Articulation*. Cambridge, Massachusetts: MIT Press.
 - Maia, M (2000). Evidentiality processes in Karaja. Published online at <http://www.museunacional.ufrj.br/linguistica/membros/maia/publicacoes.htm>
 - Wittgenstein, L., (1969). *Preliminary Studies for the 'Philosophical Investigations': Generally Known as the Blue and Brown Books*. Preface by R. Rhees. Oxford: Basil Blackwell.