

A Model of Bootstrapping for the Data-Oriented Infant

Dave Cochran,
Linguistics and English Language
School of Philosophy, Psychology and Language Sciences,
University of Edinburgh

0. Introduction

My intention in this paper is to outline and motivate a novel theory of syntactic bootstrapping, Exemplar-Based Phonological Bootstrapping (EBPB), based on Stochastic Tree-Substitution Grammar. Because I have outlined the basic structure of STSG in other places for assessed work on this degree, I will not recapitulate that work in the body of the present paper, but will instead quote it in full in Appendix A, and hereafter assume that the reader is acquainted with STSG.

My hypothesis is that the ontogenetic origin of syntactic structure lies in phonological structure. I will first outline my theory as to how this takes place, and proceed to show that the theory is motivated by the empirical research into the schedule of early phonological and syntactic development.

There is however a caveat that must be appended to all of this. STSG is a probabilistic, emergentistic theory of language, and to properly determine whether or not the theory I am about to outline really does make the predictions I think it makes, it will be necessary to model the theory computationally. The present paper, therefore, only amounts to a summary of my reasons for believing that such a computational project is worth pursuing, and likely to be successful.

1. Background

All the models of Data-Oriented Parsing (for language) developed so far operate by taking a large parsed corpus, already annotated with POS tags, and in some cases also Lexical-Functional Grammar annotations (Bod 1998, p126-43) or First Order Predicate Logic expressions (Bod, Bonnema and Scha 1996). Since the formalism is typically taken as a technology rather than a theory, this is not typically a problem, but if, as I do, you believe STSG to be the correct model for the human language faculty, the unavoidable question is, where do the trees come from? Below, I propose a model to of an answer to that question. The model takes inspiration from Bod (2002), in which it is shown that Data-Oriented Parsing can also be applied successfully to analysing music – From there follows the idea that it might also be applied to prosody, and indeed to phonological structure in general. Sadly, due to a lack of suitably phonologically annotated corpora, and indeed the lack of a generally agreed system of phonological annotation for all levels of phonological structure, this application of DOP has yet to be investigated (Rens Bod, pers. comm.). However, DOP's success in diverse cognitive modalities (music, language and scientific reasoning – see Bod (2005)) gives me reason to hope that it is likely to succeed with phonological structure too, at least once a suitable corpus is available. Figure 1 (over) shows an example tree from such a hypothetical corpus. Notice that the tree-structure consists of distinct layers, and the different node labels each can only be applied to one layer; in other words, the tree is non-recursive. What I propose is an STSG-based interpretation of Morgan & Demuth's (1996a) "phonological bootstrapping"

hypothesis, that infants acquire structured representations of ambient language phonological structure as a first step towards lexicon and and syntax.

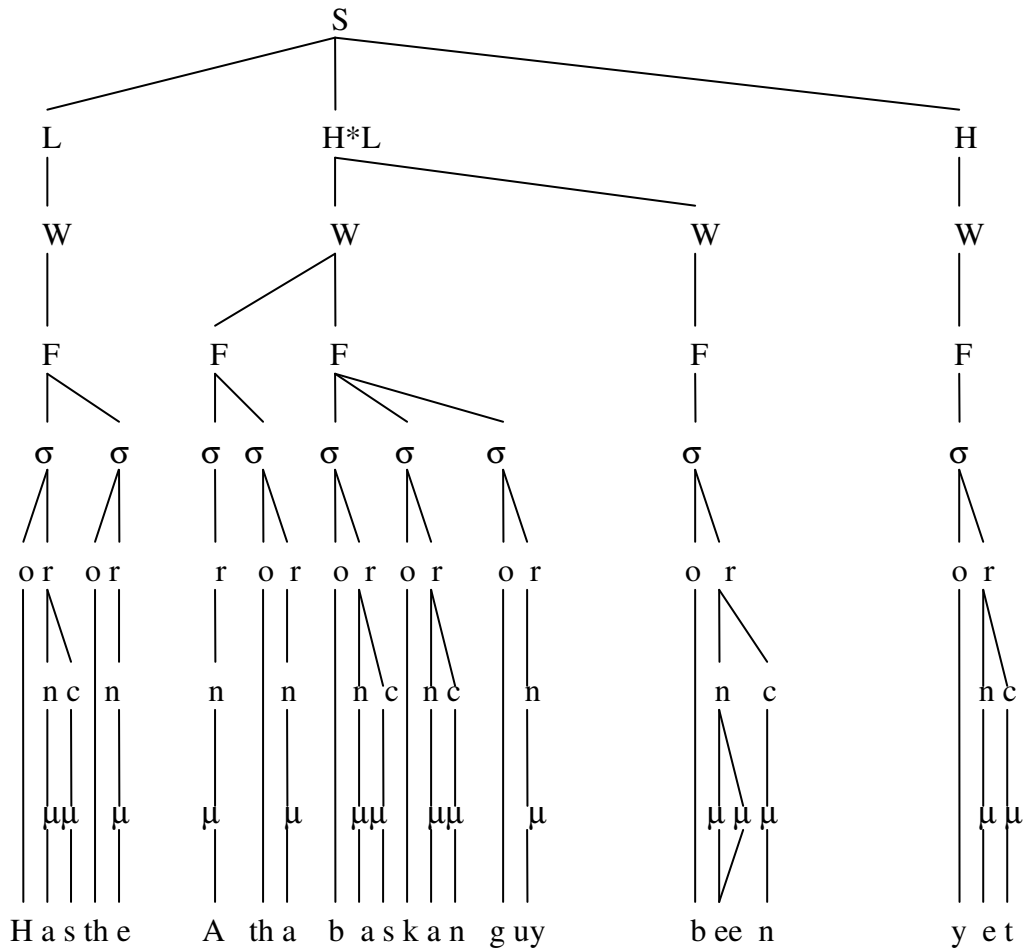


Figure 1; A phonological parse of “Has the Athabaskan guy been yet?” (Gussenhoven & Jacobs 1998, Goldsmith 1990). I have tried to follow the phonological structure of the sentence as closely as possible, ignoring syntactic boundaries.

The schedule of development I propose is as follows;

2. Phonological Parsing 1

The first simple tree structures are a product of simple collocation spotting. A pair of acoustic events (phonemes, syllables, or, in utero, pitch contours) that co-occur are represented under simple tree structures, of the following form;

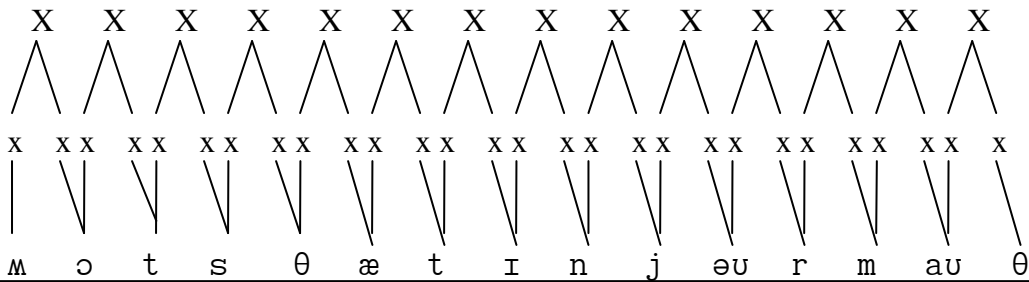


Fig 2; a simple tree for every diphone of a common parental utterance. I have used phonemes merely for diagrammatic simplicity. In fact, simple (CV) syllables are more likely to be the atomic events in such structures (see below).

The aggregation of these representations in memory, weighted for frequency, is formally isomorphic to a First-Order Hidden Markov Model (Jurafsky and Martin 2000, chapter 7), though the ordinality of the HMM is not essential to the model proposed here, nor that the states of the HMM are phonemes or syllables. That said, syllables do seem more likely; Saffran, Aslin and Newport (1996) found that 8-month old babies are sensitive to transitional probabilities between syllabic units of CV form in a prosodically undifferentiated synthesised speech-stream. What is rather more important to the theory is that these are representations in short-term memory only – most will atrophy in a short time, except those that are frequently reinforced, by which I simply mean the most frequently occurring. In fact the picture is slightly more complicated than this; Mehler, Jusczyk, Lambertz, Halsted, Bertoncini and Amiel-Tison (1988) used a high-amplitude sucking paradigm to demonstrate that neonates can distinguish their ambient language from a foreign one, but cannot tell two foreign languages apart, and can do this even when the signal is low-pass filtered, removing all segmental information. This shows that infants are sensitized to prosodic structures of their ambient language from speech signals signal low-pass filtered in utero. Therefore the disyllable (or *n*-syllable, or *n*-phone, etc) trees will be combined with a pre-existing superordinate tier of trees for prosodic structures, and segmental trees crossing prosodic constituents may be excluded. As some trees fix themselves in long-term memory and others fail to, the prosodic and segmental tree structures merge to form unified, consistent tree structures.

3. Phonological Parsing 2

More complex phonological tree-structures arise from the analysis of new inputs by the decomposition and recombination of subtrees, in the fashion of DOP. Furthermore, the tiers of the prosodic and syllabic layers are enriched with a subordinate sub-syllabic tier of layers, reflecting syllable structure as represented in fig. (1), with the onset of babbling. This takes advantage of the natural rhythms of articulatory motor gestures - “mandibular oscillations”, in MacNeilage’s (1998) terminology, which identifies the motor rhythms of the jaw of pre-speech in ontogeny with oscillations evolved in some 200 million years ago in the mammalian lineage for food processing.

I therefore hypothesise that before their first words can analyse heard utterances into phonological structures similar to fig (1)¹. The relation of these representations to babbling is a complex one, which I do not have the time or space to cover here; the empirical data points to both cross-linguistic regularities in babbling patterns, which Davis, Kern, MacNeilage, Koçbas and Zink (2005) argue is predicted by the structure and rhythm of these jaw movements, combined with a lack of independence of control over articulators. On the other hand, ambient language does seem to influence babbling patterns; for instance Boysson-Bardies and Vihman (1991) found that the frequencies of sounds varied significantly between groups of babies from French-, English-, Japanese- and Swedish-speaking homes, in ways that seemed to be influenced by the ambient language; for example, they found that French infants produced more labials in their babbling, than Swedish, who in turn produced more than American and Japanese; the same ranking as was found in their adult reference sample. To properly integrate these interactions between ambient language and physiology into the present model would require a Stochastic Tree-Substitution based model of motor memory, as advocated in Cochran (2005a), but which has yet to be developed; but in short, I would expect early babbling to begin as random motor gestures, then to be driven by the random recombination of tree-representations of motor score, then by a mixture of motor-score subtrees and phonological-parse subtrees, which, in the process are mapped to one another and eventually unified.

4. Conceptual Development

Meanwhile conceptual development proceeds, up until this point, largely independently of language. How a child segments her perceptual environment into objects, and gathers perceptions of object-tokens under object-types is a matter for another paper, for now it is only important that a child have concepts for common sensory percepts before her first words. For instance, Leslie (1988) used a Differential Looking Paradigm to show that 27 week-old infants are sensitive to the appearance of causal interaction between different coloured moving blocks on a screen, based on whether the end of the red block's motion and the onset of the green block's motion were spatially or temporally contiguous, or both, or neither. This indicated that even pre-linguistic infants have sufficient conceptual structure to be able to discriminate between events on the basis of their internal structure.

5. Emergent Semantics and the One-Word Stage

In time the child comes to form associations between percept/concepts and phonological tree-fragments that tend to co-occur in her environment. In the first instance, these associations are constrained by phonological constituency. As these semantically-enriched phonological constituents (*phonosemes*) reach a threshold frequency, they become available to DOG, and their conceptual or perceptual correlates may act as an input to production.

After a while, the child comes to recognise that some concept/percepts correspond to string fragments that are not phonological constituents; Jusczyk, Houston and Newsome (1999) present evidence to show that 7.5 month old infants from English-speaking families consistently prefer disyllabic words with strong/weak

¹ Similar, at least in terms of the richness of the information accounted for in the tree, if not the details of its structure. Certainly I do not wish to commit myself here to any particular theory of syllable structure or intonation structure.

stress patterns; when presented with sentences containing “guitar is”, they will tend to interpret “taris” as a word and not “guitar”. They go on to show that 10.5 month-olds can segment weak/strong words correctly, on the basis of more subtle cues. However, I would posit that for a child to prefer such subtle cues as phonotactics and allophony to an obvious cue like intonation, the child must have some motivation to do so – specifically, that these phonological cues would be discovered *after* the child had learnt to pick out phonologically nonconstituent segments on the basis of environmentally present “meaning” stimuli. The result is the emergence of inconsistent trees, in which parallel structures handle the phonological parse and the analysis of meaningful structures incongruous with phonological constituency

6. The Two-Word Stage

These phonologically non-constituent concept/percept-correlate fragments (which we shall hereafter call *exotaxes*, singular *exotagm*) increase in number as the child’s vocabulary grows, until it starts to happen that more than one will commonly be found in an utterance as heard and processed by the child. Collocations of exotaxes (which need not be contiguous in the speech stream) as spotted and represented under simple tree-structures similar to those which represented phoneme-collocations in fig. (2). These exotagm-pairings (*diexotaxes*) increase in frequency until a threshold is reached and they become available for generation. This correlates to the two-word stage. Where the input to generation includes semantic correlates of an exotagm and a phonoseme, the phonoseme is reanalysed as a exotagm. Bloom, Lightbown, and Hood (1975) show, using CHILDES Corpus data, that different children adopt different “strategies” at the two-word stage; of their four subjects, two adopted a “pronominal” strategy, in which a small number of pronominal forms “‘it’, ‘there’, and ‘my’” (p.19), and used them as frames for a variety of verbs, where as the other two used mostly nouns and categorised the according to certain fixed functional roles, such as “*agent, affected object, place and possessor*” (ibid.). All the children used word ordering consistent with adult usage. This would be expected if their usage was constrained by sequence-sensitive co-occurrence statistics in their input, and I would predict that the factor determining which strategy a child preferentially uses would be determined by which words are the first to be exotactically isolated from the phonological parse.

7. Recursive Syntax

Eventually diexotaxes become frequent enough to co-occur with other diexotaxes in heard utterances as processed by the child. The crucial occurrence is when two diexotaxes occur sharing a common word. These merge to form complex exotaxes, available for DOG. As more and more exotaxes are incorporated into complex trees, one may think of the emerging corpus of exotactic structures as a “notional network” of subtrees with with subtrees in common One may even hypothesise that overlapping subtrees share neural correlates for their overlapping parts, in answer to the criticism of STSG (*qua* Linguistic Theory) that it simply requires too much of past experience to be stored in memory in the long term (Simon Kirby, pers. comm.). This process of joining-up is. strikingly similar to a toy experiment by Kauffman (1995, p.55-8), presented as a model of phase-transitions. The model goes as follows; place 10,000 buttons on a tabletop, pick two buttons at random, and join them together with a thread. Repeat the last two steps many times. To begin with, you will only pick up

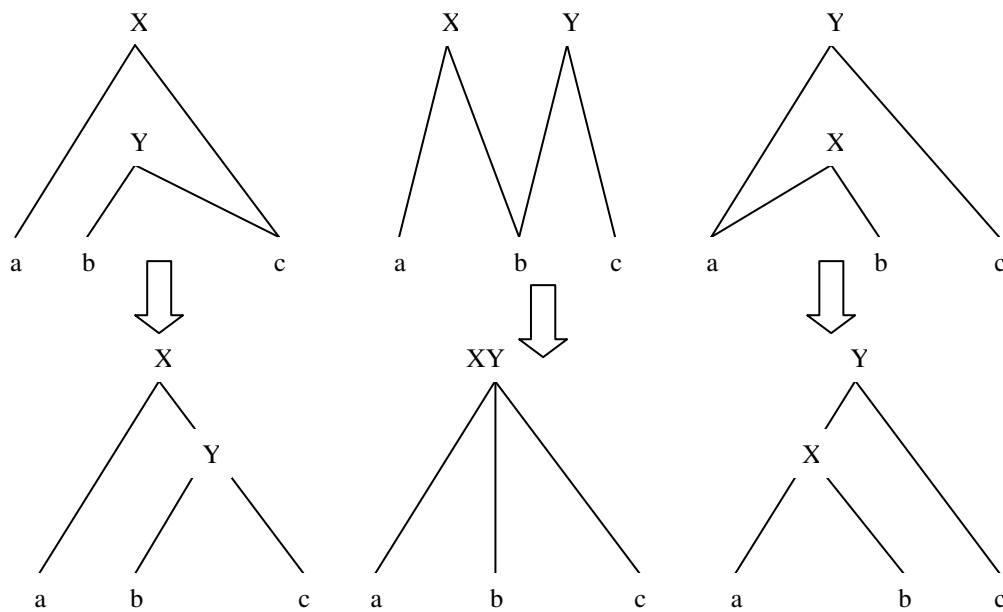


Fig 3. Three types of diexotagm merger. If we expect syntactic constituents to occur more frequently than nonconstituents, and if we expect frequent structures to act as attractors in subtree-space, merged diexotaxes representing the correct constituent structure will quickly come to dominate the child's stored representations of the language.

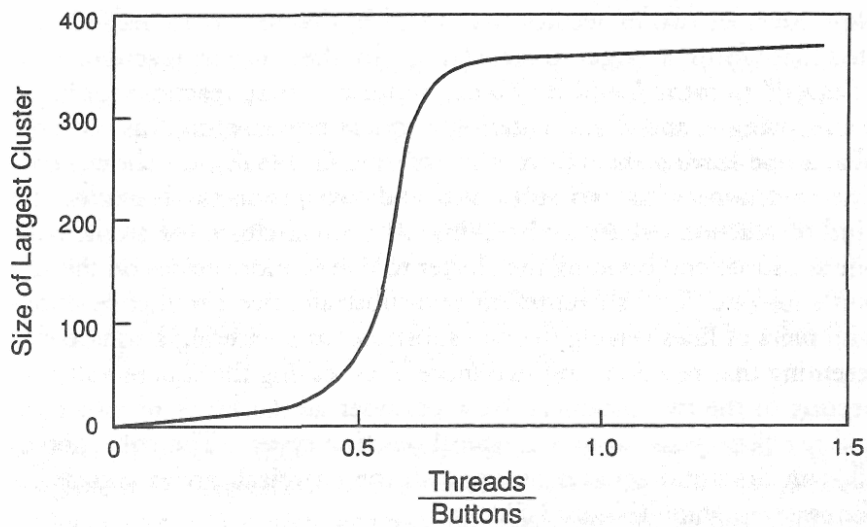


Figure 4: Threads and buttons phase transition. Note the steep sigmoidal curve when the ratio of threads over buttons approaches 0.5. Illustration from Kauffman 1995.

single buttons, but as more and more are joined, the likelihood of picking up pairs, and then larger clusters slowly increases along with the average cluster size on the table. When the ratio of threads over buttons on the table approaches 0.5, the average cluster size undergoes a sudden discontinuous jump – a phase transition “rather like separate water molecules freezing into a block of ice” (p.57). This corresponds to the emergence of multi-word utterances. Thus the parallel “exotactic” structure gradually “peels apart” from the phonological structure and becomes fully-fledged recursive syntax, leaving no attachments of meaning on the phonological structure, except;

- Intonational contours corresponding to affective states,
- Intonational contours corresponding to sentence types, such as Question Intonations, and
- Phonaesthemes.

8. Evidence from Historical Linguistics

Given the picture EBPB presents of the bootstrapping process, one could be forgiven for wondering why, in the absence of the constraint of a distinct faculty for syntax, disparities between the mapping of meaning to speech and phonological constituency exist in language at all. The answer can be found in Historical Linguistics. I appeal to the concept of *reanalysis* – that regularities in the correspondence of meaning-grounding stimuli to portions of the speech stream may arise “by accident”, without being part of the productive exemplar-base of the speaker, but be learnt as productive exemplars by subsequent generations (for example, the Middle English reanalysis of *a nadder* to *an adder* – Smith 1996) – and to the concept of *sound change* – that the phonological system of a language may undergo systematic structural change without *directly* changing other aspects of the language structure (see, eg. Kiparsky 1994). Thus, one would predict that if at any point a language did have no semantic constituents not coterminous with phonological constituents, this situation would dissipate through a slow process of entropy over several generations.

9. Conclusions

I freely admit that I have not proven this theory to be the correct account of first-language bootstrapping. That task would require years of both computational modelling, and experimental investigation; but I believe I have succeeded in showing that Exemplar-Based Phonological Bootstrapping can plausibly be expected to predict a significant number of extant experimental findings, sufficient to warrant substantial further investigation. If that investigation proves successful, then not only does language operate without genuine, output-level rules or a lexicon (Cochran 2005a, 2005b), there is no epigenetically real faculty of syntax. Although, such a thing as syntax clearly develops, it is purely a spandrel of phonology and semantics. (Gould and Lewontin 1979, for the idea of a spandrel in evolutionary biology).

References

- 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(2002): 289-308.
- Bod, R., (2005). "Towards Unifying Perception and Cognition". Prepublication.
- Bod, R., Bonnema R. and Scha, R. 1996. A Data-Oriented Approach to Semantic Interpretation. *Proceedings Workshop on Corpus-Oriented Semantic Analysis*, ECAI-96, Budapest, Hungary.
- Bloom, L., Lightbown, P. and Hood, L. (1975) *Structure and Variation in Child Language*. Monographs of the Society for Research in Child Development, serial no 160, vol 40.
- Cochran, D., (2005a), "Using Stochastic Tree-Substitution Grammar in Iterative Learning Simulations as a way of approaching issues in Diachronic Syntax", paper given at the College of Arts and Social Sciences Postgraduate Conference at the University of Aberdeen on 23rd June 2005.
- Cochran, D., (2005b), "Critically review the claim made in the statistical learning literature that no actual rules operate in the processing of language". Paper for MSc module *Psychology of Language Learning*, University of Edinburgh, 28th November 2005
- Chen, H., Cochran, D., Hanafusa, I., Laskowski, C., Ludke, K., and Ntarila, M., (2005) "Statistical Learning". Presentation for MSc module *Psychology of Language Learning*, University of Edinburgh, 22nd November 2005.
- Davis, B., Kern, S., MacNeilage, P., Koçbas, D. and Zink, I., (2005) "*Vocalization Patterns in Canonical Babbling: A cross-Linguistic Perspective*", proc. of Xth IASCL, Berlin, Germany, 25 - 29 July.
- Goldsmith, J., (1990) *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell.
- Gould, S., and Lewontin, R. (1979) "The spandrels of San Marco and the Panglossion paradigm: a critique of the adaptationist programme", *Proceedings of the Royal Society London B*. 205, pp. 581-598.
- Gussenhoven, C., & Jacobs, H., (1998) *Understanding Phonology*. London: Arnold.
- Jurafsky, D., and Martin, J. (2000) *Speech and Language Processing*. Upper Saddle River, N.J. : Prentice Hall.
- Kauffman, S. (1995). *At Home in the Universe: The Search for Laws of Self-Organization and Complexity*. New York: Oxford University Press.
- Kiparsky, P. (1994) "The phonological basis of sound change". In John A. Goldsmith, ed., *The handbook of phonological theory*, pp. 640-670.
- Leslie, A. (1988). "The necessity of illusion: Perception and thought in infancy". In L. Weiskrantz (Ed.), *Thought without language*, (pp. 185–210). Oxford: Clarendon Press/Oxford University Press.

- MacNeilage, P. (1998) "The frame/content theory of evolution of speech production". *Behavioral and Brain Sciences*, 21:499--511
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143-178.
- Morgan, J. and Demuth, K. (1996a) *Signal to syntax: bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Morgan, J. and Demuth, K. (1996b) "Signal to syntax: An Overview", in Morgan, J. and Demuth, K. 1996a, p.1-22.
- Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606-621.
- Smith, J., (1996) *An Historical Study of English: Function, Form and Change*. London : Routledge.
- Steedman M. (1996) "Phrasal Intonation and the Acquisition of Syntax", in Morgan, J. and Demuth, K. 1996a, p331-342.
- Venditti, J., Jun, S., and Beckman, M., (1996) "Prosodic Cues to Syntactic and Other Linguistic Structures in Japanese, Korean and English", in Morgan, J. and Demuth, K. 1996a, p287-312
- Xu, F., and Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology*, 30, 111-153.

Appendix A: STSG in outline

The following is not presented for assessment, but for the reader's information. The paper proper proceeds under the assumption that the reader knows at least this much about STSG. The summary is quoted from Cochran 2005b;

This summary recapitulates material in [Cochran 2005a] and Cochran *et al* 2005. The simplest manifestation of STSG is DOP1, as described in Bod 1998 (p12-23 and 40-50), though more sophisticated versions exist. The parser uses a large parsed corpus² divided into a training corpus and a smaller corpus against which the parser is tested. The parser breaks every tree in the training corpus down into all its possible subtrees, according to the wellformedness rules below.

- 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

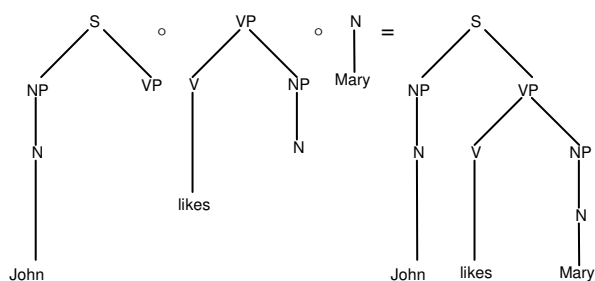


Figure 2: A derivation of "John likes Mary".

"o" is the operator for the tree-substitution operation.

The parser is given test corpus strings and builds up new parse-trees for these 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, substituting in additional subtrees, the topmost node of which must carry the same label as the node to be substituted. (see figure 2).

For each possible parse, there will be several possible derivations, and for some sentences, multiple parses may be possible. For each subtree, its probability is calculated as its total frequency of occurrence over the number of subtrees with the same root node. The probability of a derivation is the product of the probabilities of its subtrees, and the probability of a parse is the sum of the probabilities of its possible derivations. The output of the parser is most probable parse. Bod (ibid p.54) reports accuracies of 85% on the ATIS³ corpus.

² 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

³ Air Transport Information System – part of the Penn Treebank.