A Linguistic Investigation into Unsupervised DOP

Rens Bod
School of Computer Science
University of St Andrews
ILLC, University of Amsterdam
rb@cs.st-and.ac.uk

Abstract

Unsupervised Data-Oriented Parsing models (U-DOP) represent a class of structure bootstrapping models that have achieved some of the best unsupervised parsing results in the literature. While U-DOP was originally proposed as an engineering approach to language learning (Bod 2005, 2006a), it turns out that the model has a number of properties that may also be of linguistic and cognitive interest. In this paper we will focus on the original U-DOP model proposed in Bod (2005) which computes the most probable tree from among the shortest derivations of sentences. We will show that this U-DOP model can learn both rule-based and exemplar-based aspects of language, ranging from agreement and movement phenomena to discontiguous contructions, provided that productive units of arbitrary size are allowed. We argue that our results suggest a rapprochement between nativism and empiricism.

1 Introduction

This paper investigates a number of linguistic and cognitive aspects of the unsupervised data-oriented parsing framework, known as U-DOP (Bod 2005, 2006a, 2007). U-DOP is a generalization of the DOP model which was originally proposed for supervised language processing (Bod 1998). DOP produces and analyzes new sentences out of largest and most probable subtrees from previously analyzed sentences. DOP maximizes what has been called the 'structural analogy' between a sentence and a corpus of previous sentence-structures (Bod 2006b). While DOP has been successful in some areas, e.g. in ambiguity resolution, there is also a serious shortcoming to the approach: it does not account for the acquisition of *initial* structures. That is, DOP assumes that the structures of previous linguistic experiences are already given and stored in a corpus. As such, DOP can at best account for adult language use and has nothing to say about language acquisition.

In Bod (2005, 2006a), DOP was extended to unsupervised parsing in a rather straightforward way. This new model, termed U-DOP, again starts with the notion of tree. But since in language learning we do not yet know which trees should be assigned to initial sentences, it is assumed that a language learner will initially allow (implicitly) for all possible trees and let linguistic experience decide which trees are actually learned. That is, U-DOP generates a new sentence by reconstructing it out of the largest possible and most frequent subtrees from all possible (binary) trees of previous sentences. This has resulted in state-of-the-art performance for English, German and Chinese corpora (Bod 2007).

Although we do not claim that U-DOP provides any near-to-complete theory of language acquisition, we intend to show in this paper that it can learn a variety of linguistic phenomena, some of which are exemplar-based, such as idiosyncratic constructions, others of which are typically viewed as rule-based, such as auxiliary fronting and subject-verb agreement. We argue that U-DOP can be seen as a rapprochement between nativism and empiricism. In particular, we argue that there is a fallacy in the argument that for syntactic facets to be learned they must be either innate or in the input data: they can just as well emerge from an analogical process without ever hearing the particular facet and without assuming that it is hard-wired in the mind.

In the following section, we will start by reviewing the original DOP framework in Bod (1998). In section 3 we will show how DOP can be

Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition, pages 1–8,
Prague, Czech Republic, June 2007

generalized to language learning, resulting in U-DOP. Next, in section 4, we show that the approach can learn idiosyncratic constructions. In section 5 we discuss how U-DOP can learn agreement phenomena, and in section 6 we extend our argument to auxiliary movement. We end with a conclusion.

2 Review of ‘supervised’ DOP

In its original version, DOP derives new sentences by combining subtrees from previously derived sentences. One of the main motivations behind the DOP framework was to integrate rule-based and exemplar-based aspects of language processing (Bod 1998). A simple example may illustrate the approach. Consider an extremely small corpus of only two phrase-structure trees that are labeled by traditional categories, shown in figure 1.

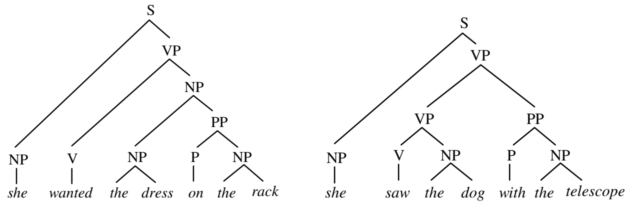


Figure 1. An extremely small corpus of two trees

A new sentence can be derived by combining subtrees from the trees in the corpus. The combination operation between subtrees is called *label substitution,* indicated as ∘. Starting out with the corpus of figure 1, for instance, the sentence *She saw the dress with the telescope* may be derived as shown in figure 2.

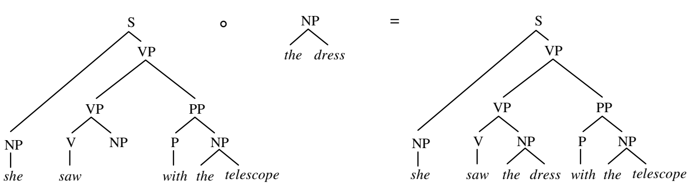


Figure 2. Analyzing a new sentence by combining subtrees from figure 1

We can also derive an alternative tree structure for this sentence, namely by combining three (rather than two) subtrees from figure 1, as shown in figure 3. We will write $(t \circ u) \circ v$ as $t \circ u \circ v$ with the convention that ∘ is *left*-associative.

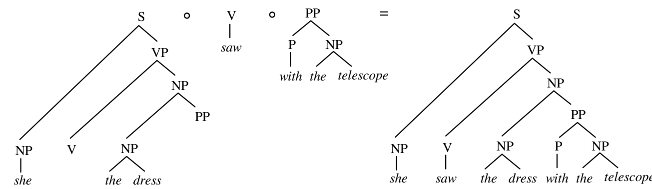


Figure 3. A different derivation for the same sentence

DOP’s subtrees can be of arbitrary size: they can range from context-free rewrite rules to entire sentence-analyses. This makes the model sensitive to multi-word units, idioms and other idiosyncratic constructions, while still maintaining full productivity. DOP is consonant with the view, as expressed by certain usage-based and constructionist accounts in linguistics, that any string of words can function as a construction (Croft 2001; Tomasello 2003; Goldberg 2006; Bybee 2006). In DOP such constructions are formalized as lexicalized subtrees, which form the productive units of a *Stochastic Tree-Substitution Grammar* or *STSG*.

Note that an unlimited number of sentences can be derived by combining subtrees from the corpus in figure 1. However, virtually every sentence generated in this way is highly ambiguous, yielding several syntactic analyses. Yet, most of these analyses do not correspond to the structure humans perceive. Initial DOP models proposed an exclusively probabilistic metric to rank different candidates, where the ‘best’ tree was computed from the frequencies of subtrees in the corpus (see Bod 1998).

While it is well known that the frequency of a structure is a very important factor in language comprehension and production (Jurafsky 2003), it is not the only factor. Discourse context, semantics and recency also play an important role. DOP can straightforwardly take into account discourse and semantic information if we have corpora with such information from which we take our subtrees, and the notion of recency can be incorporated by a frequency-adjustment function (Bod 1998). There is, however, an important other factor which does not correspond to the notion of frequency: this is the *simplicity* of a structure (cf. Frazier 1978; Chater 1999). In Bod (2002), the simplest structure was formalized by the *shortest derivation* of a sentence consisting of the fewest subtrees from the corpus. Note that the shortest derivation will include the largest possible subtrees from the corpus, thereby *maximizing the structural overlap between a sentence and previous sentence-*

structures. Only in case the shortest derivation is not unique, the frequencies of the subtrees are used to break ties among the shortest derivations. This DOP model assumes that language users maximize what has been called the *structural analogy* between a sentence and previous sentence-structures by computing the most probable tree with largest structural overlaps between a sentence and a corpus. We will use this DOP model from Bod (2002) as the basis for our investigation of unsupervised DOP.

We can illustrate DOP's notion of structural analogy with the examples given in the figures above. DOP predicts that the tree structure in figure 2 is preferred because it can be generated by just two subtrees from the corpus. Any other tree structure, such as in figure 3, would need at least three subtrees from the training set in figure 1. Note that the tree generated by the shortest derivation indeed tends to be structurally more similar to the corpus (i.e. having a larger overlap with one of the corpus trees) than the tree generated by the longer derivation. Had we restricted the subtrees to smaller sizes -- for example to depth-1 subtrees, which makes DOP equivalent to a (probabilistic) context-free grammar -- the shortest derivation would not be able to distinguish between the two trees in figures 2 and 3 as they would both be generated by 9 rewrite rules.

When the shortest derivation is not unique, we use the subtree frequencies to break ties. The 'best tree' of a sentence is defined as the most probable tree generated by a shortest derivation of the sentence, also referred to as 'MPSD'. The probability of a tree can be computed from the relative frequencies of its subtrees, and the definitions are standard for Stochastic Tree-Substitution Grammars (STSGs), see e.g. Manning and Schütze (1999) or Bod (2002). Interestingly, we will see that the exact computation of probabilities is not necessary for our arguments in this paper.

3 U-DOP: from sentences to structures

DOP can be generalized to language learning by using the same principle as before: language users maximize the structural analogy between a new sentence and previous sentence-structures by computing the most probable shortest derivation. However, in language *learning* we cannot assume that the correct phrase-structures of previously heard sentences are already known. Bod (2005) therefore proposed the following generalization of DOP, which we will simply refer to as U-DOP: *if a language learner does not know which syntactic tree should be assigned to a sentence, s/he initially allows (implicitly) for all possible trees and let linguistic experience decide which is the 'best' tree by maximizing structural analogy (i.e. the MPSD).*

Although several alternative versions of U-DOP have been proposed (e.g. Bod 2006a, 2007), we will stick to the computation of the MPSD for the current paper. Due to its use of the shortest derivation, the model's working can often be predicted without any probabilistic computations, which makes it especially apt to investigate linguistic facets such as auxiliary fronting (see section 6).

From a conceptual perspective we can distinguish three learning phases under U-DOP, which we shall discuss in more detail below.

(i) Assign all unlabeled binary trees to a set of sentences

Suppose that a language learner hears the following two ('Childes-like') sentences: *watch the dog* and *the dog barks*. How could a rational learner figure out the appropriate tree structures for these sentences? U-DOP conjectures that a learner does so by allowing any fragment of the heard sentences to form a productive unit and to try to reconstruct these sentences out of most probable shortest combinations.

Consider the set of all unlabeled binary trees for the sentences *watch the dog* and *the dog barks* given in figure 4. Each node in each tree is assigned the same category label *X*, since we do not (yet) know what label each phrase will receive.

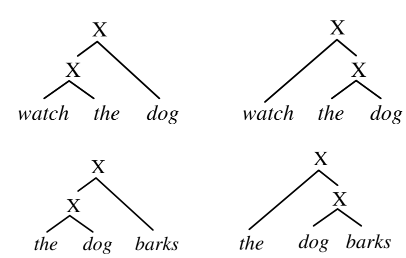


Figure 4. The unlabeled binary tree set for *watch the dog* and *the dog barks*

Although the number of possible binary trees for a sentence grows exponentially with sentence length, these binary trees can be efficiently represented by means of a chart or tabular diagram. By adding pointers between the nodes we obtain a structure known as a shared parse forest (Billot and Lang 1989).

(ii) Divide the binary trees into all subtrees

Figure 5 exhaustively lists the subtrees that can be extracted from the trees in figure 4. The first subtree in each row represents the whole sentence as a chunk, while the second and the third are 'proper' subtrees.

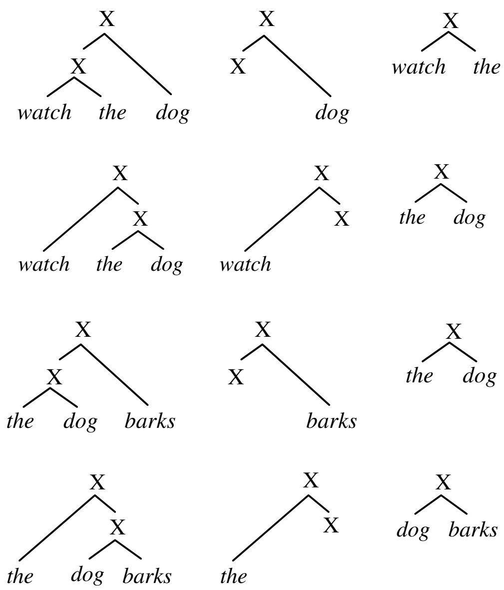


Figure 5. The subtree set for the binary trees in figure 4.

Note that while most subtrees occur once, the subtree [*the dog*]$_X$ occurs twice. There exist efficient algorithms to convert all subtrees into a compact representation (Goodman 2003) such that standard best-first parsing algorithms can be applied to the model (see Bod 2007).

(iii) Compute the 'best' tree for each sentence

Given the subtrees in figure 5, the language learner can now induce analyses for a sentence such as *the dog barks* in various ways. The phrase structure [*the* [*dog barks*]$_X$]$_X$ can be produced by two different derivations, either by selecting the large subtree that spans the whole sentence or by combining two smaller subtrees:

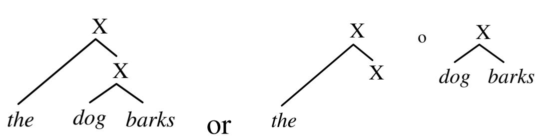


Figure 6. Deriving *the dog barks* from figure 5

Analogously, the competing phrase structure [[*the dog*]$_X$ *barks*]$_X$ can also produced by two derivations:

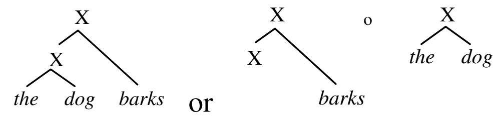


Figure 7. Other derivations for *the dog barks*

Note that the shortest derivation is not unique: the sentence *the dog barks* can be trivially parsed by any of its fully spanning trees. Such a situation does not usually occur when structures for *new* sentences are learned, i.e. when we induce structures for a held-out test set using all subtrees from all possible trees assigned to a training set. For example, the shortest derivation for the new 'sentence' *watch dog barks* is unique, given the set of subtrees in figure 5. But in the example above we need subtree frequencies to break ties, i.e. U-DOP computes the most probable tree from among the shortest derivations, the MPSD. The probability of a tree is compositionally computed from the frequencies of its subtrees, in the same way as in the supervised version of DOP (see Bod 1998, 2002). Since the subtree [*the dog*]$_X$ is the only subtree that occurs more than once, we can predict that the most probable tree corresponds to the structure [[*the dog*]$_X$ *barks*]$_X$ in figure 7 where *the dog* is a constituent. This can also be shown formally, but a precise computation is unnecessary for this example.

4 Learning constructions by U-DOP

For the sake of simplicity, we have only considered subtrees without lexical labels in the previous section. Now, if we also add an (abstract) label to each word in figure 4, then a possible subtree would the subtree in figure 9, which has a discontiguous yield *watch X dog*, and which we will therefore refer to as a "discontiguous subtree".

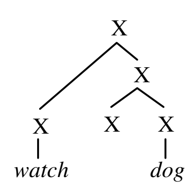


Figure 9. A discontiguous subtree

Thus lexical labels enlarge the space of dependencies covered by our subtree set. In order for U-DOP to take into account both contiguous and non-contiguous patterns, we will define the total tree-set of a sentence as the set of all unlabeled trees that are unary at the word level and binary at all higher levels.

Discontiguous subtrees, like in figure 9, are important for acquiring a variety of constructions as in (1)-(4):

(1) Show me the *nearest* airport *to* Leipzig.
(2) BA carried *more* people *than* cargo in 2005.
(3) *What is* this scratch *doing* on the table?
(4) Don't *take* him *by surprise*.

These constructions have been discussed at various places in the literature, and all of them are discontiguous in that the constructions do not appear as contiguous word strings. Instead the words are separated by 'holes' which are sometimes represented by dots as in *more … than …*, or by variables as in *What is X doing Y* (cf. Kay and Fillmore 1999). In order to capture the syntactic structure of discontiguous constructions we need a model that allows for productive units that can be partially lexicalized, such as subtrees. For example, the construction *more ... than …* in (2) can be represented by a subtree as in figure 10.

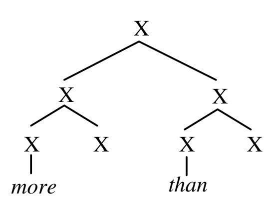


Figure 10. Discontiguous subtree for *more...than...*

U-DOP can learn the structure in figure 10 from a few sentences only, using the mechanism described in section 3. While we will go into the details of learning discontiguous subtrees in section 5, it is easy to see that U-DOP will prefer the structure in figure 10 over a structure where e.g. [*X than*] forms a constituent. First note that the substring *more X* can occur at the end of a sentence (in e.g. *Can I have more milk?)*, whereas the substring *X than* cannot occur at the end of a sentence. This means that [*more X*] will be preferred as a constituent in [*more X than X*]. The same is the case for *than X* in e.g. *A is cheaper than B*. Thus both [*more X*] and [*than X*] can appear separately from the construction and will win out in frequency, which means that U-DOP will learn the structure in figure 10 for the construction *more … than …*.

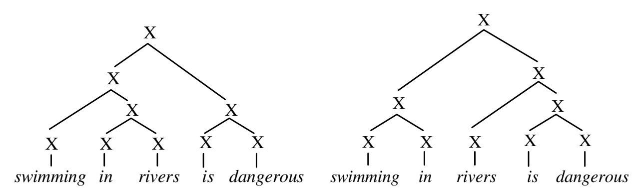
Once it is learned, (supervised) DOP enforces the application of the subtree in figure 10 whenever a new form using the construction *more ... than ...* is perceived or produced because the particular subtree will directly cover it and lead to the shortest derivation.

5 Learning agreement by U-DOP

Discontiguous context is important not only for learning constructions but also for learning various syntactic regularities. Consider the following sentence (5):

(5) Swimming in rivers is dangerous

How can U-DOP deal with the fact that human language learners will perceive an agreement relation between *swimming* and *is*, and not between *rivers* and *is*? We will rephrase this question as follows: which sentences must be perceived such that U-DOP can assign as the best structure for *swimming in rivers is dangerous* the tree 16(a) which attaches the constituent *is dangerous* to *swimming in rivers*, and not an incorrect tree like 16(b) which attaches *is dangerous* to *rivers*? Note that tree (a) correctly represents the dependency between *swimming* and *is dangerous*, while tree (b) misrepresents a dependency between *rivers* and *is dangerous*.



(a) (b)

Figure 16. Two possible trees for *Swimming in rivers is dangerous*

It turns out that we need to observe only one additional sentence to overrule tree (b), i.e. sentence (6):

(6) Swimming together is fun

The word *together* can be attached either to *swimming* or to *is fun*, as illustrated respectively by 17(a) and 17(b) (of course, *together* can also be attached to *is* alone, and the resulting phrase *together is* to *fun*, but our argument still remains valid):

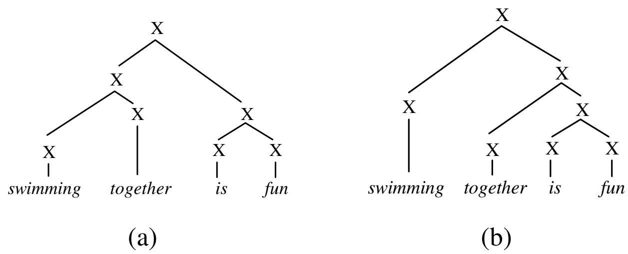


Figure 17. Two possible trees for *Swimming together is fun*

First note that there is a large common subtree between 16(a) and 17(a), as shown in figure 18.

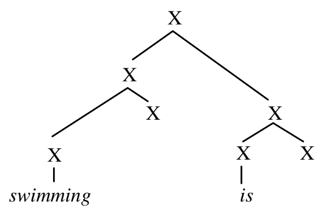


Figure 18. Common subtree in the trees 16(a) and 17(a)

Next note that there is not such a large common subtree between 16(b) and 17(b). Since the shortest derivation is not unique (as both trees can be produced by directly using the largest tree from the binary tree set), we must rely on the frequencies of the subtrees. It is easy to see that the trees 16(a) and 17(a) will overrule respectively 16(b) and 17(b), because 16(a) and 17(a) share the largest subtree. Although 16(b) and 17(b) also share subtrees, they cover a smaller part of the sentence than does the subtree in figure 18. Next note that for every common subtree between 16(a) and 17(a) there exists a corresponding common subtree between 16(b) and 17(b) except for the common subtree in figure 18 (and one of its sub-subtrees by abstracting from *swimming*). Since the frequencies of all subtrees of a tree contribute to its probability, if follows that figure 18 will be part of the most probable tree, and thus 16(a) and 17(a) will overrule respectively 16(b) and 17(b).

However, our argument is not yet complete: we have not yet ruled out another possible analysis for *swimming in rivers is dangerous* where *in rivers* forms a constituent together with *is dangerous*. Interestingly, it suffices to perceive a sentence like (7): *He likes swimming in river.* The occurrence of *swimming in rivers* at the end of this sentence will lead to a preference for 16(a) because it will get a higher frequency as a group. An implementation of U-DOP confirmed our informal argument.

We conclude that U-DOP only needs three sentences to learn the correct tree structure displaying the dependency between the subject *swimming* and the verb *is*, known otherwise as "agreement". Once we have learned the correct structure for subject-verb agreement by the subtree in figure 18, (U-)DOP enforces agreement by the shortest derivation.

It can also be shown that U-DOP still learns the correct agreement if sentences with incorrect agreement, like **Swimming in rivers are dangerous*, are heard as long as the *correct* agreement has a higher frequency than the incorrect agreement during the learning process.

6 Learning 'movement' by U-DOP

We now come to what is often assumed to be the greatest challenge for models of language learning, and what Crain (1991) calls the "parade case of an innate constraint": the problem of auxiliary movement, also known as auxiliary fronting or inversion. Let's start with the typical examples, which are similar to those used in Crain (1991), MacWhinney (2005), Clark and Eyraud (2006) and many others:

(8) The man is hungry

If we turn sentence (8) into a (polar) interrogative, the auxiliary *is* is fronted, resulting in sentence (9).

(9) Is the man hungry?

A language learner might derive from these two sentences that the first occurring auxiliary is fronted. However, when the sentence also contains a relative clause with an auxiliary *is*, it should not be the first occurrence of *is* that is fronted but the one in the main clause, as shown in sentences (11) and (12).

(11) The man who is eating is hungry
(12) Is the man who is eating hungry?

There is no reason that children should favor the correct auxiliary fronting. Yet children do produce the correct sentences of the form (12) and rarely if ever of the form (13) even if they have not heard the correct form before (see Crain and Nakayama 1987).

(13) *Is the man who eating is hungry?

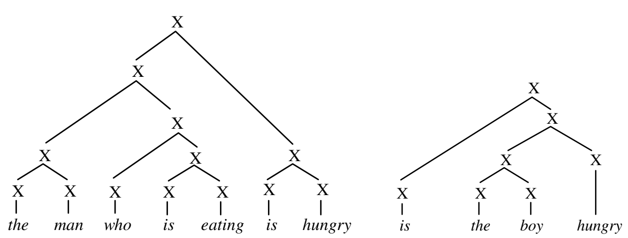
How can we account for this phenomenon? According to the nativist view, sentences of the type in (12) are so rare that children must have innately specified knowledge that allows them to learn this facet of language without ever having seen it (Crain and Nakayama 1987). On the other hand, it has been claimed that this type of sentence is not rare at all and can thus be learned from experience (Pullum and Scholz 2002). We will not enter the controversy on this issue, but believe that both viewpoints overlook a very important alternative possibility, namely that auxiliary fronting needs neither be innate nor in the input data to be learned, but may simply be an emergent property of the underlying model.

How does (U-)DOP account for this phenomenon? We will show that the learning of auxiliary fronting can proceed with only *two* sentences:

(14) The man who is eating is hungry
(15) Is the boy hungry?

Note that these sentences do not contain an example of the fact that an auxiliary should be fronted from the main clause rather than from the relative clause.

For reasons of space, we will have to skip the induction of the tree structures for (14) and (15), which can be derived from a total of six sentences using similar reasoning as in section 5, and which are given in figure 20a,b (see Bod forthcoming, for more details and a demonstration that the induction of these two tree structures is robust).



(a) (b)

Figure 20. Tree structures for *the man who is eating is hungry* and *is the boy hungry?* learned by U-DOP

Given the trees in figure 20, we can now easily show that U-DOP's shortest derivation produces the correct auxiliary fronting, without relying on any probability calculations. That is, in order to produce the correct interrogative, *Is the man who is eating hungry*, we only need to combine the following two subtrees from the acquired structures in figure 20, as shown in figure 21 (note that the first subtree is discontiguous):

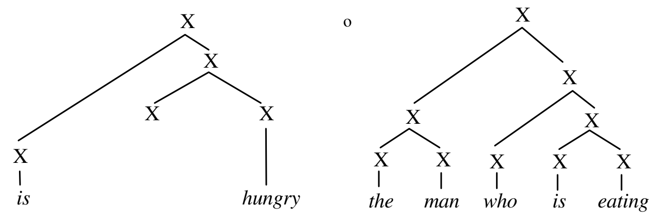


Figure 21. Producing the correct auxiliary fronting by combining two subtrees from figure 20

On the other hand, to produce the sentence with incorrect auxiliary fronting **Is the man who eating is hungry?* we need to combine many more subtrees from figure 20. Clearly the derivation in figure 21 is the shortest one and produces the correct sentence, thereby blocking the incorrect form.[1]

Thus the phenomenon of auxiliary fronting needs neither be innate nor in the input data to be learned. By using the notion of shortest derivation, auxiliary fronting can be learned from just a couple sentences only. Arguments about frequency and "poverty of the stimulus" (cf. Crain 1991; MacWhinney 2005) are therefore irrelevant – provided that we allow our productive units to be of arbitrary size. (Moreover, learning may be further eased once the syntactic categories have been induced. Although we do not go into category induction in the current paper, once unlabeled structures have been found, category learning turns out to be a relatively easy problem).

Auxiliary fronting has been previously dealt with in other probabilistic models of structure learning. Perfors et al. (2006) show that Bayesian model selection can choose the right grammar for auxiliary fronting. Yet, their problem is different in that Perfors et al. start from a set of given grammars from which their selection model has to choose the correct one. Our approach is more congenial to Clark and Eyraud (2006) who show that by distributional analysis in the vein of Harris (1954) auxiliary fronting can be correctly predicted. However, different from Clark and Eyraud, we have shown that U-DOP can also learn complex, discontiguous constructions. In order to learn both rule-based phenomena like auxiliary inversion and exemplar-based phenomena like idiosyncratic constructions, we believe we need

[1] We are implicitly assuming here an extension of DOP which computes the most probable shortest derivation given a certain meaning to be conveyed. This semantic DOP model was worked out in Bonnema et al. (1997) where the meaning of a sentence was represented by its logical form.

the richness of a probabilistic tree grammar rather than a probabilistic context-free grammar.

7 Conclusion

We have shown that various syntactic phenomena can be learned by a model that only assumes (1) the notion of recursive tree structure, and (2) an analogical matching algorithm which reconstructs a new sentence out of largest and most frequent fragments from previous sentences. The major difference between our model and other computational learning models (such as Klein and Manning 2005 or Clark and Eyraud 2006) is that *we start with trees*. But since we do not know which trees are correct, we initially allow for all of them and let analogy decide. Thus we assume that the language faculty (or 'Universal Grammar') has knowledge about the notion of tree structure but no more than that. Although we do not claim that we have developed any near-to-complete theory of all language acquisition, our argument to use only recursive structure as the core of language knowledge has a surprising precursor. Hauser, Chomksy and Fitch (2002) claim that the core language faculty comprises just 'recursion' and nothing else. If one takes this idea seriously, then U-DOP is probably the first fully computational model that instantiates it: U-DOP's trees encode the ultimate notion of recursion where every label can be recursively substituted for any other label. All else is analogy.

References

Billot, S. and B. Lang, 1989. The Structure of Shared Forests in Ambiguous Parsing. *Proceedings ACL 1989*.

Bod, R. 1998. *Beyond Grammar*. Stanford: CSLI Publications.

Bod, R. 2002. A Unified Model of Structural Organization in Language and Music, *Journal of Artificial Intelligence Research*, 17, 289-308.

Bod, R. 2005. Combining Supervised and Unsupervised Natural Language Processing. *The 16th Meeting of Computational Linguistics in the Netherlands (CLIN 2005)*.

Bod, R. 2006a. An All-Subtrees Approach to Unsupervised Parsing. *Proceedings ACL-COLING 2006*, 865-872.

Bod, R. 2006b. Exemplar-Based Syntax: How to Get Productivity from Examples. *The Linguistic Review* 23, 291-320.

Bod, 2007. Is the End of Supervised Parsing in Sight?. *Proceedings ACL 2007*, Prague.

Bod, forthcoming. From Exemplar to Grammar: How Analogy Guides Language Acquisition. In J. Blevins and J. Blevins (eds.) *Analogy in Grammar*, Oxford University Press.

Bonnema, R., R. Bod and R. Scha, 1997. A DOP Model for Semantic Interpretation. *Proceedings ACL/EACL 1997*, Madrid, Spain, 159-167.

Bybee, J. 2006. From Usage to Grammar: The Mind's Response to Repetition. *Language* 82(4), 711-733.

Chater, N. 1999. The Search for Simplicity: A Fundamental Cognitive Principle? *The Quarterly Journal of Experimental Psychology*, 52A(2), 273-302.

Clark, A. and R. Eyraud, 2006. Learning Auxiliary Fronting with Grammatical Inference. *Proceedings CONLL 2006*, New York.

Crain, S. 1991. Language Acquisition in the Absence of Experience. *Behavorial and Brain Sciences* 14, 597-612.

Crain, S. and M. Nakayama, 1987. Structure Dependence in Grammar Formation. *Language* 63, 522-543.

Croft, B. 2001. *Radical Construction Grammar*. Oxford University Press.

Frazier, L. 1978. On Comprehending Sentences: Syntactic Parsing Strategies. PhD. Thesis, U. of Connecticut.

Goldberg, A. 2006. Constructions at Work: the nature of generalization in language. Oxford University Press.

Goodman, J. 2003. Efficient algorithms for the DOP model. In R. Bod, R. Scha and K. Sima'an (eds.). *Data-Oriented Parsing*, CSLI Publications, 125-146.

Harris, Z. 1954. Distributional Structure. *Word* 10, 146-162.

Hauser, M., N. Chomsky and T. Fitch, 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?, *Science* 298, 1569-1579.

Jurafsky, D. 2003. Probabilistic Modeling in Psycholinguistics. In Bod, R., J. Hay and S. Jannedy (eds.), *Probabilistic Linguistics*, The MIT Press, 39-96.

Kay, P. and C. Fillmore 1999. Grammatical constructions and linguistic generalizations: the What's X doing Y? construction. *Language*, 75, 1-33.

Klein, D. and C. Manning 2005. Natural language grammar induction with a generative constituent-context model. *Pattern Recognition* 38, 1407-1419.

MacWhinney, B. 2005. Item-based Constructions and the Logical Problem. *Proceedings of the Second Workshop on Psychocomputational Models of Human Language Acquisition*, Ann Arbor.

Manning, C. and H. Schütze 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press.

Perfors, A., Tenenbaum, J., Regier, T. 2006. Poverty of the Stimulus? A rational approach. *Proceedings 28th Annual Conference of the Cognitive Science Society*. Vancouver

Pullum, G. and B. Scholz 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19, 9-50.

Tomasello, M. 2003. *Constructing a Language*. Harvard University Press.