The importance of rule restrictions for CCG according to Kuhlmann et al. and its relevance to the debate about the generative capacity of natural language

Logical methods in natural language production – Myrthe Tielman – 3345343

Introduction

The debate about the generative capacity of natural languages goes back to the work of Chomsky who claimed that natural languages were not finite state (Chomsky, 1956). In the 1980's, research turned to the question if context-free languages were powerful enough to describe all natural language phenomena. The main problems for context-free languages are cross-serial dependencies in Dutch and Swiss German and unbounded scrambling phenomena found in German and Korean. An example of the kind of cross-serial dependency in Dutch of which context-free grammar has a problem is found in 1.

1. *… dat Jan Piet Marie de kinderen zag helpen leren zwemmen*
   …that Jan Piet Marie the children saw help teach swim
   …that Jan saw Piet help Marie teach the children to swim

In generative grammars, these cross-serial dependencies are formalised, usually by assigning a string letter to each pair of dependent words. For the sentence in 1. this would yield the string abcdabcd. In the remainder of this paper I will use these strings as representations of languages. Research has shown that context-free grammars do not have the power to assign the correct structural descriptions to these kinds of sentences (Bresnan et al., 1982). The question that arose from this discovery, though, is to what extent natural languages are beyond context-free? We know that they are richer than context-free grammar, but not as powerful as context-sensitive grammar. To describe the class natural languages should be found in Joshi formalised the concept of mildly context-sensitive grammar with the definition found in 2. (Joshi, 1985).

2. *A set L of languages is mildly context-sensitive iff*
   a. *L contains all context-free languages*
   b. *L can describe cross-serial dependencies: There is an $n \geq 2$ such that $\{w^k \mid w \in T^*\} \in L$ for all $k \leq n$.*
   c. *The languages in L are polynomially parsable, i.e., $L \subset$ PTIME.*
   d. *The languages in L have the constant growth property.*

   A formalism F is mildly context-sensitive iff the set $\{L \mid L = L(G)$ for some $G \in F\}$ is mildly context-sensitive.

Many different mildly context-sensitive formalisms have been developed and not all have the same generative capacity. Although there is a consensus on the idea that natural languages are mildly context-sensitive, which formalisms approach the theoretical idea of mildly context-sensitivity best is still question for debate. Although the answer by no means agreed upon, many contributions to this debate have been made in the past decades. One recent contribution to this debate is the work of Kuhlmann et al. (2010) on the importance of rule restrictions in CCG. In this paper, I will give an overview the work of Kuhlmann et al on CCG and the proofs they present. I will then give an overview of the debate on the generative capacity of natural language as it stands today and the relevance of the work of Kuhlmann et al. herein.

CCG

Combinatory Categorial Grammar (CCG) was first developed by Steedman (2000) with the goal to provide a "principled theory of natural grammar more directly compatible on the one hand with certain syntactic phenomena that flagrantly disrupt order and constituency, including coordination, extraction, and intonational phrasing, and on the other with psychological and computational mechanisms that can map such surface forms onto interpretable meaning representations“[1] (Steedman, 2000). CCG is a mild context-sensitive language and thus obeys the principles described by Joshi (1985). The attribute which makes that CCS is mildly context-sensitive is its flexible composition rules, which enable the grammar to generate those word orders that are beyond context-free. It has been shown, however, that pure first-order CCG’s (which can use generalized composition rules only and does not allow restrictions of the instances of these rules) cannot generate $a^nb^nc^n$, making its generative capacity strictly smaller than other mildly-context sensitive formalisms such as TAG (Koller & Kuhlmann, 2009). This is a problem for pure first-order CCG in the way that it means that it is not a fully lexicalized formalism as it has to allow grammar-specific rules to be able to generate patterns found in natural language. The ideal, however, is to have a lexicalized language which can deal with all natural language phenomena, because this would give us a universal set of grammatical rules applicable to all natural languages. The question posed by Koller & Kuhlmann (2009) is if this problem also applies to pure CCG with higher-order categories.

This question is addressed by Kuhlmann et al. in a later paper (2010). In this section an overview of their argument will be given, as well as a discussion of the contribution of their outcome to the debate on the generative capacity needed to model natural languages. Kuhlmann et al. begin their argument by defining pure CCG, a CCG which is lexicalized. A pure combinatory categorical grammar (PCCG) is a construct $G = (A, \Sigma, L, s)$ where A is an alphabet of atomic categories, $s \in A$ a distinguished atomic category called the *final category*, Σ a finite set of *terminal symbols* and L a finite relation between symbols in Σ and categories over A called the *lexicon*. The elements of L are called *lexicon entries* and are represented with the notation $\sigma \vdash x$, where $\sigma \in \Sigma$ and x is a category over A. Categories occurring in a lexicon entry are called *lexical* categories. Derivations in a PCCG grammar G is represented as a derivation tree which is constructed by lining up the lexical categories of the symbols in the given string and applying the rules from Figure 1 to adjacent pairs until a single category remains.

$x/y \;\; y \Rightarrow x$	forward application	$>$
$y \;\; x\backslash y \Rightarrow x$	backward application	$<$
$x/y \;\; y/z \Rightarrow x/z$	forward harmonic composition	$>B$
$y\backslash z \;\; x\backslash y \Rightarrow x\backslash z$	backward harmonic composition	$<B$
$x/y \;\; y\backslash z \Rightarrow x\backslash z$	forward crossed composition	$>B_{\times}$
$y/z \;\; x\backslash y \Rightarrow x/z$	backward crossed composition	$<B_{\times}$

Figure 1.

Several restrictions are often applied to CCG. One restriction employed by Kuhlmann et al. is degree restriction, stating an upper bound for the degree of composition rules which can be used in derivations of $n \geq 0$. The second kinds of restrictions are the rule restrictions, restricting which rules may be used in which cases. These restrictions eliminate the lexicalized property of CCG, but are used in most practical CCG grammars. Kuhlmann et al. show that PCCG is still more expressive than context-free grammar. More relevant though, they also prove that the weak generative capacity of a CCG with rule restrictions is strictly greater than the generative capacity of PCCG. In the proof for this theorem the notions of active and inactive arguments are adopted. An argument is active if it ends in a primary premise, being matched against a subcategory of the secondary premise. It is inactive if it ends in a secondary premise, being consumed as part of a higher order-argument.[2] Kuhlmann et al. now define the core problem PCCG’s have with generating $a^nb^nc^n$. All PCCG’s which are able to

[1] Steedman 2000, “The Syntactic Process”, page xi.
[2] Kuhlmann et al 2010, page 538

generate $a^nb^nc^n$ are also able to generate strings not of the form $a^nb^nc^n$, the problem is thus that they overgenerate. This overgeneration is caused by the property of PCCG that arguments can be saturated prematurely. In terms of active and inactive arguments, PCCG allows the saturation of almost all active arguments of a category before it is used as a secondary premise and no more than one active argument must be transferred to the conclusion of that premise. Kuhlmann et al. now turn to show that any derivation containing a category with at least two active argument can be transformed into a different derivation where the arguments can be saturated prematurely. To characterize this transformation, the set of rewriting rules in figure 2. are introduced.

$$\frac{\dfrac{{}^{A}\,x/y \quad {}^{B}\,y\beta/z}{x\beta/z} \quad {}^{C}\,z\gamma}{x\beta\gamma} \xRightarrow{\text{R1}} \frac{x/y \quad \dfrac{y\beta/z \quad z\gamma}{y\beta\gamma}}{x\beta\gamma} \qquad \frac{\dfrac{{}^{B}\,y\beta/z \quad {}^{A}\,x\backslash y}{x\beta/z} \quad {}^{C}\,z\gamma}{x\beta\gamma} \xRightarrow{\text{R2}} \frac{\dfrac{y\beta/z \quad z\gamma}{y\beta\gamma} \quad x\backslash y}{x\beta\gamma}$$

$$\frac{{}^{C}\,z\gamma \quad \dfrac{{}^{A}\,x/y \quad {}^{B}\,y\beta\backslash z}{x\beta\backslash z}}{x\beta\gamma} \xRightarrow{\text{R3}} \frac{x/y \quad \dfrac{z\gamma \quad y\beta\backslash z}{y\beta\gamma}}{x\beta\gamma} \qquad \frac{{}^{C}\,z\gamma \quad \dfrac{{}^{B}\,y\beta\backslash z \quad {}^{A}\,x\backslash y}{x\beta\backslash z}}{x\beta\gamma} \xRightarrow{\text{R4}} \frac{\dfrac{z\gamma \quad y\beta\backslash z}{y\beta\gamma} \quad x\backslash y}{x\beta\gamma}$$

Figure 2. (γ represents sequence of arguments (possibly empty), β represents sequence of arguments where the first argument is active)

These transformation rules can be applied to the grandparent of any highest critical node. Here a node u is the highest critical node if its corresponding category is the secondary premise of a rule and contains more than one active argument and there is no other node with these previous properties with a shorter distance to the root than u. Kuhlmann et al. show that all rewritings of a derivation are finite and that no rewriting rule increase the degree of composition operations in the proof. The transformed derivations have several interesting properties. Kuhlmann et al. prove that the set of all transformed derivations of a grammar give a context-free language. From this result, they show that every language generated by a PCCG has a Parikh-equivalent context-free sublanguage. This in turn indicates that a PCCG cannot generate all languages that can be generated with CCG, since CCG can generate $a^nb^nc^n$ and the only Parikh-equivalent sublanguage of $a^nb^nc^n$ is $a^nb^nc^n$ itself.

The question Kuhlmann et al. now turn to is if this also applies to multi-model CCG (MM-CCG). MM-CCG is kind of extension of CCG which introduces the concept of slash types into CCG. Slash types work in the way that they specify which rules can apply to which arguments by requiring that the slash type of the rule and the argument be the same. The most common type of MM-CCG was developed by Baldridge and Kruijff (2003) and will be referred to as B&K-CCG. B&K-CCG impose a hierarchy of four slash types ranging from most general to most specific. These slash types are applied to the combinatory ways as seen in figure 3.

$x/_{\star}y \quad y \Rightarrow x$	forward application
$y \quad x\backslash_{\star}y \Rightarrow x$	backward application
$x/_{\diamond}y \quad y/_{\diamond}z\beta \Rightarrow x/_{\diamond}z\beta$	forward harmonic composition
$x/_{\times}y \quad y\backslash_{\times}z\beta \Rightarrow x\backslash_{\times}z\beta$	forward crossed composition
$y\backslash_{\diamond}z\beta \quad x\backslash_{\diamond}y \Rightarrow x\backslash_{\diamond}z\beta$	backward harmonic composition
$y/_{\times}z\beta \quad x\backslash_{\times}y \Rightarrow x/_{\times}z\beta$	backward crossed composition

Figure 3.

With these new rules, properties which in CCG had to be expressed in the form of rule restrictions can be expressed in the types of lexical items. This is a great advantage as it allows pure B&K-CCG to express certain language-specific properties that pure CCG cannot. Kuhlmann et all, however, show that even though B&K-CCG offers slash types as a way to specify language specific properties in the lexicon, it still needs rule restrictions to generate $a^n b^n c^n$. The point in the proof which is different for B&K-CCG are the transformation rules. Kuhlmann et al. prove that even with the slash-types grammatical re-writing rules can be given. This means that pure B&K-CCG still has a strictly smaller weak generative capacity than B&K-CCG which allows rule restrictions.

These finds show that the weak generative capacity of CCG depends crucially on rule restriction and that CCG's with this generative capacity can therefore not be fully lexicalized, even when slash types are introduced.

Parsing complexity of Natural Language

The question to answer is now, what is the relevance of the discovery that CCG's with the generative capacity to model natural languages cannot be fully lexicalized? The first step in answering this question is to take a closer look at the debate on the generative capacity needed for natural language as it stands today. There is a general consensus on the idea that a generative grammar for natural languages should be mildly-context sensitive, so somewhere between context-free and context-sensitive grammar and having the described in 2. The problem, however, is that many mildly context-sensitive formalisms exist and that they do not all have the same generative capacity.

Within the mildly context-sensitive formalisms we can distinguish two important categories of formalisms with a different generative capacity. The category with formalisms most powerful formalisms includes LCFRS, MCFG, simple RCG, MG, set-local MCTAG and finite-copying LFG. The mildly context-sensitive formalism which is least powerful TAG, LIG, tree-MCTAC, EPDA and CCG. There is no settled answer as to which of these two categories is best for modelling natural language, but there seemed to be a consensus that the larger class would be most suited, mostly because of it being very robust. Recent work has raised the question, however, if the answer shouldn't be sought somewhere in between. Kanazawa (2009) introduced the idea that well-nested versions of formalisms provide a better approximation of the notion of mildly context-sensitive as introduced by Joshi (1985). This notion of well-nestedness is a restriction on languages and states that pairs of disjoint dependency trees should not cross. Kanazawa argues that the category including well-nested MCFG's (which lies between the two categories earlier defined in generative capacity) should be preferred above the category including MCFG's.

The question now is in what way the findings from Kuhlmann et al. are relevant for this debate. Kuhlmann et al. make no explicit claims about their views, but the results they present do have some important consequences. The first is that although (multi-model) CCG has been proven useful in modelling human languages in the past, it is theoretically not ideal because it is not fully lexicalized. This can be interpreted as an indication that we should let go of the ideal of a lexicalized formalism and that a non-lexicalized formalism may simply be better at generating natural languages. Another view is to take the results to indicate that CCG is, from a theoretical viewpoint, not a formalism we want for modelling natural languages. For CCG we can now say that it is either not lexicalized or not mildly context-sensitive. Seeing how we would like a formalism to be as close to mildly context-sensitive as possible when modelling natural language, we turn to non-lexicalized CCG. This property of not being lexicalized does, however, implicate that if CCG is a good model for natural language, natural language is not lexicalized. This in turn would mean that languages do not differ solely on a lexical level with a set of universal grammar rules. If natural language was not lexicalized, languages would differ in respect to the grammatical rules as well. This, however, is a very controversial notion which should not be adopted easily. For these reasons the conclusion seems to be that CCG is simply not a suited formalism for modelling natural language. The next question to be asked be if this negative result for CCG transcends to the equivalent formalisms and if this can be seen as indication that a formalism with a larger generative capacity such as well-nested MCFG should be preferred when modelling natural language. This, however, is not an easy question to be answered and requires more in-depth work on several mildly context-sensitive formalisms.

Conclusion

In this paper I have presented the results from Kuhlmann et al. on the importance of rule restrictions in CCG. Kuhlmann et al. have shown that these rule restrictions are essential for the generative capacity of CCG being equivalent to TAG. This importance of rule restrictions shows us that a lexicalized CCG has a weaker generative capacity than mildly context sensitive formalisms. In the debate about the generative capacity of natural language the main question today is what class of mildly context-sensitive formalisms approaches Joshi's original definition of mild context-sensitivity best. The results from Kuhlmann et al. show us that CCG is only mildly context-sensitive when non-lexicalized, which can be seen as an indication that CCG is not a theoretically suited formalism to model natural language.

References

Bresnan, J. , Kaplan, R.M., Peters, S., Zaenen, A. **Cross-Serial Dependencies in Dutch,** *Linguistic Inquiry* , Vol. 13, No. 4 (Autumn, 1982), pp. 613-635

Chomsky, N. **Three models for the description of language**, IRE Transactions on Information Theory, 2:113-124 (1956)

Joshi, A.K.: **Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions**? In: Dowty, D.R., Karttunen, L., Zwicky, A.M. (eds.) *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, pp. 206–250. Cambridge University Press, Cambridge (1985)

Kallmeyer, L. **Parsing Beyond Context-Free Grammars**, Springer (2010)

Kanazawa, M.: **The convergence of well-nested mildly context-sensitive grammar formalisms**. In: *An invited talk given at the 14th Conference on Formal Grammar*, Bordeaux, France (July 2009), http://research.nii.ac.jp/~kanazawa/

Koller, A., Kuhlmann,M. **Dependency trees and the strong generative capacity of CCG**. In *Proceedings of the Twelfth Conference of the European Chapter of the Association for Computational Linguistics (EACL*), pages 460–468, Athens, Greece.(2009)

Kuhlmann, M., Koller, A., Satta, G., **The importance of Rule Restrictions in CCG**, In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 534–543.Uppsala, Sweden, (2010)

Shieber, S.M. **Evidence against the Context-Freeness of Natural Language,** *Linguistics and Philosophy* , Vol. 8, No. 3 (Aug., 1985), pp. 333-343

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