# Kracht/Sportiche 214, Winter 2003: Combinatory Categorial Grammar — Discussion

### 1. Agreement

CCG offers no principled solution to agreement. The agreement symbols are added as diacritics. However, Steedman assumes that CCG has a unification component. Categories are thus not primitive, but are in turn made from attribute value statements, say [NUM : plu]. No precise definitions are given, though.

## 2. Case and Type Raising

In CCG, type is raising is free, but one can only raise using special types. Thus, the set of  $\tau$  such that  $\alpha$  may be raised to  $\tau/(\tau \setminus \alpha)$  (or  $\tau \setminus (\tau/\alpha)$ ) is parametrically restricted. This picture is somewhat simplistic. It would allow an NP to raise to  $S/(S \setminus NP)$  (subject) or to  $(S \setminus NP) \setminus ((S \setminus NP)/NP)$ (object). In case marking languages, only nominative marked NPs may raise to S/(S NP), and only accusative marked NPs to (S NP) ((S NP)/NP). Steedman drops a few remarks on Page 45 on this topic, saying that there is a correlation between case marking and type raising, but the matter is left unresolved. In view of the fact that many languages have NP-internal agreement, one take on the matter would be to install agreement features into the basic categories (using AV-pairs as indicated above) and to parametrize the raising potential of NPs to their case, so that nominative NPs can only raise to become subjects, accusative NPs only to become objects. This means that case is a lexical feature, while function is a grammatical one. On the other hand, Ps selecting case present a difficulty. The NP must raise before it combines with the P head. It will then have the category

since an accusative assigning P head has category PP/NP[*acc*]. In constructions with a verb selecting a PP–complement, the NP will for example assume the category

 $(S\NP)\((S\NP)/NP[acc])$ 

For the verb has category  $((S \setminus NP)/PP)$  and so may forward compose with the P head. Notice that this correctly discloses the case marking properties

of the P. The construction is reminiscent of an often made proposal that verbs incorporate the P heads (or at least some of them).

# 3. Multiple Specification of Word Order and Parametric Licensing

Word order is typically specified twice in categorial grammars. First, it is specified in the verb, and then in the category of the raised NPs. For example, a verb initial language has transitive verbs of category (S/NP)/NP, while a verb medial language has transitive verbs of category (S/NP)/NP. On the other hand, if an object raises over the verb, it assumes the category

$$(S/NP) \setminus ((S/NP)/NP)$$

while in a verb medial language it assumes the category

### $(S\NP)\((S\NP)/NP)$

This seems like an unnecessary iteration of the facts. However, in CCG this gets exploited in a rather nonstandard way. Namely, there are verb medial languages (Dutch, German) in which both leftward and rightward gapping are grammatical, while at least in theory only one possibility should exist (the one that exists in English).

- (1) ..., dass Karl Paul ein Radio gab und Hans Peter das Buch.
- (2) ..., dass Karl Paul ein Radio und Hans Peter das Buch gab.
- (3) ... that Karl gave Paul a radio and Hans Peter the book.

(4)  $*\ldots$  that Karl Paul a radio and Hans gave Peter the book.

Now, one type of gapping is natural, namely backward gapping, since the verb is looking for the object to its left. The other one is found only in SVO or VSO languages. Notice now that although the subordinate clause in German and Dutch is strictly verb final, the main clause shows V2. Steedman interprets this as follows. The verbs of Dutch also allow for a main clause VSO order. Indeed, questions have type VSO (At Jan appels? ('Did Jan eat apples?')). To account for the fact that there can be one argument preceding the verb (like relative pronouns), these elements may receive the following category:

(a) 
$$(S_{-sub}^{+CP}/(S_{-sub} \setminus NP))$$
  
(b)  $(S_{-sub}^{+CP}/(S_{-sub}/NP))$   
(c)  $((S_{-sub}^{+CP}/\alpha)/((S_{-sub}/\alpha) \setminus NP))$ 

where  $S_{-sub}^{+CP}$  is the category of a main clause CP with topicalization,  $S_{-sub}$  the category of a main CP without topicalization.

Thus, Dutch and German are treated as VSO languages with an extraposed (= topicalized) constituent, thus accounting for the fact that the nature of this constituent is quite free. The conjunct revealing rule for Dutch (and presumably also German) is the same as for English:

Virtual conjunct revealing rule.

$$\alpha \quad \rightsquigarrow \quad \beta \quad \alpha \backslash \beta$$

where  $\beta = S/\$$ .

What we still need to understand is why this does not allow for SOV+SO gapping in a strict SOV language. The crucial factor is the *parametric licensing* of type raising. Steedman says the following:

The restriction limits  $\tau \setminus \alpha$  and  $\tau / \alpha$  to types that are permitted under the (informally defined) Principle of Categorial Type Transparency. Among other things I will assume it prevents infinite recursion of type-raising [...]. The restriction [...] also means that, for example in English, as opposed to German,  $\tau \setminus \alpha$  cannot be instantiated as (S\NP)\NP. At least in English, and possibly in all languages, we can assume that this restriction limits  $\tau \setminus \alpha$ to a finite set of categories.

The idea is that if there is no verb that projects SVO or VSO order, there will be no type raising instance that produces NPs looking to the left for the verb. But the virtual conjunct revealing rule requires the formation of a category  $\alpha \setminus (S/\$)$ . Apparently, Japanese conforms to this prediction. In Japanese, a subject NP and a object NP can be forward composed:

$$\frac{\text{Ken-ga}}{\text{S}/(\text{S}\setminus\text{NP}[nom])} \quad (\text{S}\setminus\text{NP}[nom]) \setminus ((\text{S}\setminus\text{NP}[nom])/\text{NP}[acc])} \\ \frac{\text{S}/((\text{S}\setminus\text{NP}[nom]) \setminus \text{NP}[acc])}{\text{S}/((\text{S}\setminus\text{NP}[nom]) \setminus \text{NP}[acc])}$$

The result is, however, not of the required form. Since there is no verb that has either SVO or VSO order, the NPs cannot be type raised in such a way that backward gapping becomes possible.

Similarly, in English leftward gapping is impossible, since the composition of subject and object yields a constituent of category S((S/NP)/NP). This

is of the required kind, but leftward application is still not permitted. The only possibility is to apply coordination to the lefthand disclosed constituent, but that can recompose only to the right.



## 4. Semantics

The conjunct revealing rule may be syntactically powerful, but it compromises the architecture of the grammar.

Virtual conjunct revealing rule.

$$\alpha \quad \rightsquigarrow \quad \beta \quad \alpha \backslash \beta$$

where  $\beta = S/\$$ .

The problem with this rule is that the semantic analysis cannot be undone, since there are no functions that can achieve this. There have been proposals to use higher–order unification (Shieber, and Dalrymple and Shieber). However, HOU is a rather powerful device. Instead, Steedman assumes that the functions producing ex post analysis already exist.

Virtual conjunct revealing rule.

$$\alpha: N \quad \rightsquigarrow \quad \beta: \theta'' N \quad \alpha \backslash \beta: \lambda y. N$$

where  $\beta = S/\$$ .

Here,  $\theta''N$  is the theme of N. It is thus predicted that the gap is always a theme, and that the semantics of  $\beta$  only depends on semantics plus the thematic structure of the first conjunct. Steedman quotes (Kuno, 1976) in support of this. Kuno's FSP Principle of Gapping is as follows.

a. Constituents deleted by gapping must be contextually known. On the other hand, the two constituents left behind by gapping by gapping necessarily represent new information and, therefore, must be paired with constituents in the first conjunct that represent new information.

- b. It is generally the case that the closer a given constituent is to the sentence final position, the newer information it represents in the sentence.
- c. Constituents that are clearly marked for nonanaphoricity necessarily represent new information in violation with (b). Similarly, constituents that appear closest to sentence final position necessarily represent old information (in violation of (b)) if coreferential constituents appear in the corresponding preceding discourse.

Kuno argues in his article against the structure based approaches to gapping by Hankamer and others. In (Kornfilt, 2001) it is claimed that the postverbal elements in a Turkish sentence present old, presupposed information. Thus, it is predicted that one cannot have SVO+SO gapping, because that would mean that O is new, not old.

# 5. Gapping and its Kin

#### Some Terminological remarks

A number of different constructions involve coordination in one way or another.

(5)	(Coordination)
	John gave Mary a tulip and Mary a rose.
(6)	(Right node raising)
	John was reading and Mary only filing this article.
(7)	(VP ellipsis)
	John is singing and Mary, too.
(8)	(Stripping)
	John is eating bread in the bed and candies, too.
(9)	(Gapping)
	John gave Mary a tulip and Peter Susan a rose.
(10)	(Sluicing)
	Someone ate my bread but I don't know who.

The terminology is as follows. Coordination is only of constituents, where constituents are not believed to be as flexible as in CCG. In English, we have the structure (S(VO)), so (6) is not an instance of coordination. If the coordinated parts are left contituent parts, as in (6), we speak of right node raising; if it is a left constituent part, we speak of VP-ellipsis, since what

has been elided must be a VP. If it is an interior part of the constituent, we speak of stripping. Sluicing is of a different kind, however. Notice that stripping is the shadow image of gapping: you strip the left and right part of the interior: xyzandy' is stripping, xyzandx'z' is an instance of gapping. This inventory is however not entirely complete. Dutch and German allow quite complex patterns:

(11) Jan gab Lisa gestern eine Rose und Peter Linda eine Tulpe.Jan gave Lisa yesterday a rose and Peter Linda a tulip

This instantiates vwxyzundv'x'z'. Steedman calls this discontinuous gapping.

(12) Jacob heeft appels gegeten en Hendrik peren. Jacob has appels eaten and Hendrik pears

#### Gapping in CCG

One of the claims of CCG is that all the above with the exception of Sluicing are in fact simply different kinds of coordinations. This is principally due to the fact that CCG allows for a more flexible notion of constituent. However, it does not support a wrapping rule, as does (Dowty, 1996).

#### Problems

In (Eisenberg, 1973), Eisenberg noted that forward and backward deletion obey different criteria in German.

- (13) \*..., dass ich ein Buch und Maria einen Roman liest. ..., that I a book and Mary a novel read-PRES.3
- (14) ..., dass ich ein Buch und Maria einen Roman las. ..., that I a book and Mary a novel read-PST.1/3
- (15) ..., dass ich ein Buch lese und Maria einen Roman. ..., that I a book read-PRES.1 and Mary a novel
- (16) ..., dass ich ein Buch las und Maria einen Roman. ..., that I a book read-PST.1/3 and Mary a novel

Backward deletion is possible only in presence of identity of surface form, while forward deletion is possible also when the surface forms do not match. If they match, it is not necessary that the two are equal word forms (see (14)). Eisenberg also gives examples from coordination

- (17) \*der Antrag des oder der Lehrer the proposal the-SG or the-sc plu teacher-PL.GEN
- (18) der Antrag des oder der Dozenten the proposal the-SG or the-sc plu teacher-SG/PL.GEN

The requirement seems to be less strict in case mismatch (mit oder ohne Kinder (with or without children), where mit requires Kindern). Notice that backward deletion will also require certain further identity (say, it must be the same root that is involved), otherwise the following would be possible:

(19) This definition has one and this cat four claws.

Here is now the crux. If forward and backward gapping were subject only to the combinatorial possibilities, we would be forced to either ignore overt morphology (in the case of forward gapping) or to insist on it (in the case of backward gapping). A possible remedy for German is the following observation: only rightward gapping requires the use of the virtual conjunct revealing rule. The leftward gapping construction does not use that rule, it is constituent coordination. So, we might propose that coordination requires PF identity of the material that is 'deleted'. This is problematic at least for NP-coordination:

(20) Jan und ich tranken Bier. Jan and I drank beer

Thus, NP-coordination is exempt from strict identity, if the view is upheld that it is the same as backward deletion (for such a view see (Wilder, 1995)). Notice similar problems with tense agreement:

- (21) \*dass ich gestern einen Roman las und morgen ein Lehrbuch.
- (22) dass ich heute einen Roman lese und morgen ein Lehrbuch.
- (23) \*dass ich gestern einen Roman und morgen ein Lehrbuch lese.
- (24) dass ich heute einen Roman und morgen ein Lehrbuch lese.

Gapping is possible when the verbform is the same (present and future) but not when they are different. Curiously, however, it does not matter in which direction we gap.

(Bozşahin, 2000) notices that forward and backward gapping in Turkish are different in that only forward gapping allows for a different word order, backward gapping does not.

- (25) Hasan karides-i ye-di, istiridye-yi de Mehmet. Hasan shrimp-ACC eat, oyster-ACC and Mehmet
- (26) karides-i Hasan ye-di, Mehmet de istiridye-yi. shrimp-ACC Hasan eat-PAST, Mehmet and oyster-ACC
- (27) \*Hasan karides-i, istiridye-yi de Mehmet ye-di. Hasam shrimp-ACC, oyster-ACC and Mehmet eat
- (28) \*karides-i Hasan, Mehmet te istiridye-yi ye-di. shrimp-ACC Hasan, Mehmet and oyster-ACC eat

So, one can have SOV+OS, OSV+SO but neither SO+OSV nor OS+SOV. Thus, a similar requirement of identity is at work in Turkish, though it affects word order rather than agreement. (Kornfilt, 2001) reports, though, that the same order effects as found in German are also found in Turkish.

In languages were word order is more free (Latin) or which do not show constituency (many Australian languages), the combinatorial approach faces severe problems. We have mentioned also discontinuous constituency, whose existence Steedman acknowledges without offering an account. It has been proposed in (Hoffman, 1995) to separate the categorial spine from the alignment. This however compromises the nature of CCG, which is built on strict adjacency. Notice that Turkish, while having a basic SOV word order, is like Latin in allowing various alternative word orders. (Kornfilt, 1997) says that in coordination the NPs must be elided from the second conjunct, while the verb may or may not be elided in the first. However, the preferred option is (31) rather than (30).

- (29) Hasan ıstakoz-u pişir-di, Ali de ye-di Hasan lobster-ACC cook-PAST, Ali -and eat-PAST
- (30) Hasan ıstakoz-u pişir-di, Ali de balığ-ı Hasan lobster-ACC cook-PAST, Ali -and fish-ACC
- (31) Hasan ıstakoz-u, Ali de pişir-di Hasan lobster-ACC, Ali -and eat-PAST

Pulte noted in (Pulte, 1974) that Quechua counterexemplifies the predictions concerning the directionality of gapping. Although Quechua prefers SOV, it allows for all six permutations of (32) ((Weber, 1989)).

(32) Hwan Tumas-ta maqa-n. John Tom-OBL hit-3

Quechua allows all variants of rightward gapping, but no leftward gapping:

- (33) SVO+SO, SVO+OS, SOV+SO, SOV+OS, VSO+SO, VSO+OS VOS+SO, VOS+OS, OVS+SO, OVS+OS
- (34) \*SO+SVO, \*OS+SVO, ...

Cherokee has SOV, OVS, SVO, again no leftward gapping, but the following forward gapping patterns (again, see (Pulte, 1974)):

(35) SOV+SO, OVS+SO, SVO+SO

#### Gapping as Coordination

Standard accounts of gapping view it either ATB-extraction or as ellipsis. A third option, occasionally considered, is that in gapping the remainder actually *is a constituent* so that it is like coordination (cf. Wilder, (Zoerner, 1995)). Evidence that some instances of gapping are coordination and not ellipsis has been put forward by (Kazenin, 2001b). In Dargwa, an SOV language (ergative, with class marking), the verb agrees with the absolutive. (The transliteration is the one I found in the source. I have not been able to improve on it.)

 (36) dul mutal ma1Ha1[Qalalij (\*b-)/w-ata1Ribda.
 I.ERG Mutal.ABS.CL:1 to.Makhachkala (\*CL:1.PL)CL:1.SG-send.PAST I sent Mutal to Makhachkala.

In gapping, plural agreement is mandatory:

(37) dul mutal dil rasul ma1Ha1[Qalalij \*w-/b-ata1Ribda. I.ERG Mutal.ABS.CL:1 you Rasul.ABS.CL:1 to.Makhachkala (\*CL:1.SG)CL:1.PL-send.PAST I sent Mutal and you Rasul to Makhachkala.

Similarly for a structure  $S(O_1+O_2)V$ . A different pattern is attested in Russian. When the verb is left peripheral it must take plural agreement.

(38) Zavtra poedut/\*poedet: Kolja v Moskvu, a Vasja v Peterburg. Tomorrow, Kolja will.go.PL/\*will.go.SG to Moscow, Vasja to St. Petersburg.

In (Kazenin, 2001a), Kazenin shows that in Chuvash, leftward and rightward gapping have different properties. While leftward gapping looks very much like coordination (it optionally triggers plural agreement), this is not so for

forward gapping. Also, leftward gapping can eat into an NP, forward gapping cannot.

- (39) Vasja vyrAsla petja cAvaSla kEneke-ne il-c-E. Vasja Russian Pete Chuvash book-ACC buy-PST-3
- (40) \*Vasja vyrAsla kEneke-ne il-c-E, petja cAvaSla. Vasja Russian book-ACC buy-PST-3, Petja Chuvash
- (41) \*Vasja vyrAsla kEneke-ne il-c-E, petja cAvaSla kEneke-ne. Vasja Russian book-ACC buy-PST-3, Petja Chuvash book-ACC

#### Eating into an NP

We notice here a problem with the type raising rule in connection with gapping. It is noted that gapping may 'eat into' an NP:

(42) Die linksrheinischen Örter haben warmes Wetter und die the left-of-the-Rhine places have warm weather and the rechtsrheinischen Regen. right-of-the-Rhine rain

If this is to be interpreted as coordination, we must assume that the partial NP that lacks a noun can combine with a VP. However, this assumes that it has the category of a type raised NP lacking the head noun. However, although the virtual conjunct revealing rule would allow for this in principle, it is not clear that the required categories can be formed, for lack of parametrized categories. Notice, namely, that type raising applies only after the formation of the NP, so that it forms a 'bracket'. One could argue, however, that discontinuous NPs exist in German, revealing the possibility of raising of the left half of an NP as well as the right half. The exact details need to be worked out. Notice that one can create 'crossed' NP eatings:

(42) Die rechtsrheinischen haben deutsche Ortsnamen, die linksthe right-of-the-Rhine have German place-names, the leftrheinischen Dörfer französische. of-the-Rhine villages French

(Notice that German allows also to exchange SO with OS in both parts.)

# Conclusion

- Leftward and Forward gapping are different. Typically, leftward gapping is more heavily restricted. It may require phonetic identity of the elided part, it may require identity in word order of parallel elements, and it may disallow 'eating into an NP'.
- CCG allows for rightward gapping in SOV languages only if there are main clause verbs that have SVO or VSO order. The mechanism for rightward gapping is, however, markedly different in that it requires ex post analysis. This may well explain why the identity requirements are relaxed.
- Ex post analysis is possible in a categorial setting (i. e. assuming compositionality) only if the semantics has similar functions. CCG assumes that gapping material must be new, and so can be recovered from the first conjunct even when it is fully assembled. This does not seem to hold for Turkish, though.
- There is a distinction between gapping and coordination in that the former can trigger agreement only with the element in the clause, while in coordination agreement is with the 'conjoined' arguments. While the distinction is clear within one language, it must be noted though that also coordination may be subject to constraints that may blur the distinction (why can I not say ??Du und Peter gehst/geht/gehen in den Garten.) and that agreement may be only with one conjunct (Latin, Arabic).

# References

- (Bozşahin, 2000) C. Bozşahin. Directionality and the Lexicon: Evidence from Gapping. Unpublished Manuscript, 2000.
- (Dowty, 1996) David Dowty. Nonconstituent Coordination, Wrapping and Multimodal Categorial Grammars: Syntactic Form as Logical Form. In Kees Doets, editor, Proceedings of the Tenth International Congress of Logic, Methodology and the Philosophy of Science, 1995, Amsterdam, 1996. North-Holland.

- (Eisenberg, 1973) Peter Eisenberg. A Note on 'Identity of Constituents'. Linguistic Inquiry, 4:417 – 420, 1973.
- (Hoffman, 1995) Beryl Hoffman. The Computational Analysis of the Syntax and Semantics of "Free" Word Order in Turkish. PhD thesis, University of Pennsylvania, 1995.
- (Kazenin, 2001a) Konstantin I. Kazenin. Can gapping always be constituent coordination? Manuscript, University of Moscow/Tübingen, 2001.
- (Kazenin, 2001b) Konstantin I. Kazenin. Gapping and some agreement puzzles. Manuscript, University of Moscow/Tübingen, 2001.
- (Kornfilt, 1997) Jaklin Kornfilt. *Turkish.* Routledge, London and New York, 1997.
- (Kornfilt, 2001) Jaklin Kornfilt. Directionality of Identical Verb Deletion in Turkish Coordination. Presented at the Jorge Hankamer WebFest, 2001.
- (Kuno, 1976) Susumo Kuno. Gapping: A Functional Analysis. *Linguistic* Inquiry, 7:300 – 318, 1976.
- (Pulte, 1974) William Pulte. A Note on Gapping. Linguistic Inquiry, 4:100 - 101, 1974.
- (Weber, 1989) David John Weber. A Grammar of Hullaga (Huánoco) Quechua. Number 112 in Linguistics. University of California Press, 1989.
- (Wilder, 1995) Chris Wilder. Some properties of ellipsis in coordination. Technical Report 2, Department of Linguistics, University of Geneva, 1995.
- (Zoerner, 1995) Cyril Edward Zoerner. *Coordination: The Syntax of &P*. PhD thesis, University of California, Irvine, 1995.