Reevaluating Gapping in CCG: Evidence from English and Chinese

Nathaniel Little
Yale University
P.O. Box 200209
New Haven, CT 06520
nathaniel.little@aya.yale.edu

Abstract

This paper considers gapping data through the lens of combinatory categorial grammar (CCG) as developed in Steedman (1990, 2000). It analyzes CCG's predictive power in managing a wide variety of cross-clausal gapping data. CCG predicts the typing of the rightmost subject in cross-clausal gapping data as an object; evidence from Case supports this hypothesis. Reflexive binding in cross-clausal structures favors the Szabolcsi (1989) binding proposal, in which binding occurs at the level of the surface structure. Additionally, facts from Chinese buttress the CCG analysis, as its NP category-assignment offers a natural explanation for the ungrammaticality of gapping sentences containing non-quantified NP objects: they are unable to undergo type-shifting.

1 Setting the stage

Ross (1967) gave the name *gapping* to the following phenomenon:

(1) Harry eats beans, and Fred, potatoes.

Since Ross (1967), gapping has received varied accounts; I will center the present discussion on the theoretical problems posed by attempting to describe the data in the context of combinatory categorial grammar (CCG) as developed in Steedman (1990, 2000).

The paper will proceed as follows. First, I review the basic facts for which any theory of gapping must provide an account. Next, I describe the basic mechanics of gapping in CCG. I continue on to discuss data that proves problematic for the CCG account of gapping; I focus on the means by which CCG is able to predict many of the initially problematic data, thereby lending support for an intermediate theory of gapping – one that combines a syntactic and a semantic account. Further support for the CCG analysis comes in the form of apparent instances of gapping in Chinese, as CCG offers a principled account of the alternations between sentences containing quantified and non-quantified NP objects.

2 Gapping: preliminary facts

To begin, let us define gapping as a construction involving (at least) two similar clauses that surface in a contrastive relationship. In (1), we see a contrast between the left conjunct *Harry eats beans* and the right one, *Fred, potatoes*. The two conjuncts are joined through coordination; many theories take coordination to be a foundational property of gapping, but the story is not so simple. English gapping, for example, also occurs in comparatives:

(2) Harry eats more beans than Fred, potatoes.

It is clear from (2) that gapping is not limited to syntactic coordination. Even gapping sentences with coordination need not be marked through the presence of an overt coordinator:

(3) Some ate natto; others, rice.

Cases like (2-3) show that attempts to consider gapping as a uniquely syntactic phenomenon involving the overtly marked coordination of constituents will inevitably fall flat.

Gapping generally involves only one gap, and the item that is "gapped" is at least the main verb, if not additional material:

(4) Harry eats beans, and Fred (eats) potatoes.

The content of the additional material may vary; in many cases, it is part of the verb phrase:

(5) John bought a book at the store, and Bill, online.

The above example shows that not only is the main verb *bought* gapped, but the object *a book* is gapped, as well. In most cases, there may be only one gap – sentences containing multiple gaps are often ungrammatical:

(6) *I gave Mary a flower yesterday, and you, Bill, today.

In the above example, there are two gaps: first, the main verb *gave* is deleted, followed by a second, discontinuous gap in which the object *a flower* is deleted. As shown in (5), it is possible for both a verb and an object to be deleted (or "gapped"), so the problem is not the fact that

both items are deleted, but that they are deleted discontinuously. One gap is composed of the main verb *gave*, and the other, discontinuous gap contains the NP *a flower*. The dative NP *Bill* intervenes between the two gaps, which results in ungrammaticality. Yet if the dative object surfaces in a prepositional phrase in a right peripheral position in the clause, the sentence is grammatical, and its meaning is preserved:

(7) I gave a book to Mary yesterday, and
you, to Bill, three weeks ago.

The contrast between (6) and (7) points to a preference for placing focused material at peripheral positions within the clause. Thus, discontinuous gaps are not generally permitted either because they delete focused elements or they fail to delete non-focused material.

The disparity in grammaticality of (6-7) hints at another general property of gapping: the items in the gapped conjunct must be tied to some material that is sufficiently salient in the discourse to deliver an apparent contrast between the two (or more) conjuncts. This property is represented below (with (1) reprinted as (8a)):

(8) a. Harry eats beans, and Fred, potatoes.
b. *A* eats *B*, and *C*, *D*.

Schematically, we see the contrastive relationship set up between the pairs *A*,*B* and *C*,*D*, in which the one element of each pair is contrasted with the corresponding element of the other; that is, *A* contrasts with *C*, and *B* with *D*. This relationship must be made clear within the discourse, and sentences uttered in contexts in which such a contrast is absent will be ungrammatical.

These facts are ones that any theory of gapping must be able to explain. What follows is a discussion of CCG's ability to account for this data and for other, similar gapping phenomena.

3 Gapping in CCG

Gapping viewed through the lens of CCG offers an intermediate stage between strongly syntactic and strongly semantic accounts of gapping. The syntactic categories assigned to each lexical item reflect the notion of the gap, and a semantic focus constraint serves to limit gapping to sufficiently salient discourse contexts.

In CCG, there is no underlying syntactic structure, or in fact any deleted structure at all. CCG carries a comparatively free notion of constituency, which allows for surface-level combination between string-adjacent elements into increasingly large, concatenated constituents. In some sense, CCG's idea of constituency, which allows even (traditionally "discontinuous") elements like *Fred, potatoes* to be considered constituents, resembles Ross's (1970) proposal that gapping and VP-ellipsis target and then elide "context variables" that range over strings regardless of constituency. In CCG, the decomposition of elements in the left conjunct allows us to pick out the verb and identify a non-continuous string in the left conjunct, which may then combine with a non-continuous string in the right conjunct to derive a licit sentence. The intuition that gapping targets strings of lexical items that in many cases are either discontinuous or non-standard constituents is one easily captured by CCG; its lexical category assignment, combined with its finite set of combinatory rules, permits the derivation of non-standard constituents without appealing to other levels of structure. Thus, even though Ross's (1970) proposal deals with strings that are deleted – and CCG lacks any notion of underlying structure – the shared intuition is one of non-standard constituency, which may be targeted in instances of VP-ellipsis and (important for present purposes) gapping.

Steedman's (1990, 2000) CCG account of gapping relies on a notion of constituency that is fundamentally different from that of abstract approaches. In CCG, a constituent refers to any entity within the grammar that fulfills two criteria: it must be interpretable, and grammatical rules must be allowed to operate on it (see Steedman 1990 for further discussion). CCG assigns a category to each lexical item; a combinatory rule operates on a pair of string-adjacent lexical items. The derivability of a sentence is determined by the categories of the lexical items and their (in)ability to combine according to CCG's finite set of combinatory rules. One basic rule is function application:

(9) Function Application (> or <)
a. $X/Y \quad Y \quad \rightarrow \quad X$
b. $Y \quad X\backslash Y \quad \rightarrow \quad X$

X and *Y* may be thought of as variables corresponding to categories; directionality of the function is indicated by the direction of the slash (a forward-slash is right-looking, and a backward-slash is left-looking). Function application allows string-adjacent lexical items of the appropriate type to combine. Other rules are necessary for the derivation of sentences

containing non-traditional constituents (e.g. *cooked* and *might eat*). Function composition allows such combination:

(10) Forward Composition (>B)

a. $X/Y \quad Y/Z \rightarrow_{\mathbf{B}} X/Z$

Without forward composition, the sentence *I cooked, and might eat, the beans* would be underivable as a complete sentence:

```
(11) a. I     cooked,      and   might        eat,      the beans
        NP    (S\NP)/NP   conj  (S\NP)/VP    VP/NP     NP
                                             -------------------->
                                                      VP
     ------------------------>   ------------------------->
              S\NP                         S\NP
            --------------------->&
               [S\NP]&
               -------------------------------->
                          S\NP

     b. I     cooked,      and   might        eat,      the beans
        NP    (S\NP)/NP   conj  (S\NP)/VP    VP/NP     NP
                                -------------------------->B
                                      (S\NP)/NP
                          ---------------------------->&
                              [(S\NP)/NP]&
              ---------------------------------------<&
                    (S\NP)/NP
                    ------------------------------------------------------>
                          S\NP
        ---------------------------------------<
                    S
```

In (11b) above, the non-traditional constituent *might eat* may be derived through forward composition. Similarly, in CCG any item – word, phrase or non-canonical combination of words – may rightly be considered a constituent. Thus, as Steedman notes, a string like *Mary might* is as much a constituent as the predicate *eat the cake* would be in many abstract accounts. This relaxed notion of constituency is articulated in CCG's rules, which allow the concatenation of words into such "non-standard" constituents – including, crucially, the subject/object pairs found in the derivation of gapping sentences (whose analysis is outlined below). Because the second conjunct in gapping sentences is a constituent, coordination may apply to it. Steedman's formulation of CCG thus maintains the idea that gapping respects constituency, by loosening constituency's requirements.

Three additional rules – type-raising, forward mixing composition and decomposition – make possible the derivation of gapping sentences. I will offer a brief treatment of each, though the reader is directed to Steedman (1990, 2000) for a fuller discussion.

Type-raising, along with forward mixing composition, is necessary in order to combine the subject/object remnants in the right conjunct of gapping sentences. Type-raising turns arguments of functions into functions-over-such-functions-over-arguments (e.g. one normally thinks of nouns as arguments of verbs; in CCG, a type-raised noun becomes a function taking a verb as its argument). A general type-raising schema is shown below:

(12) Type-Raising:

$A \rightarrow_{\mathbf{T}} B/(B\backslash A)$ *or* $B\backslash(B/A)$

When applied to subject *NP*s, type-raising is instantiated as follows:

(13) Subject Type-Raising (>T)

$NP \rightarrow_{\mathbf{T}} S/(S\backslash NP)$

As *NP* objects of a transitive verb, the type-raised category for English objects is that of a function taking a transitive verb as its input and returning a one-argument verb:

(14) Object Type-Raising (<T)

$NP \rightarrow_{\mathbf{T}} (S\backslash NP)\backslash((S\backslash NP)/NP)$

Type-raising, like other rules in CCG, allows for greater combinatory possibilities and is necessary for the derivation of gapping sentences (among other phenomena). Taking our simple gapping sentence, type-raising gives us the following:

```
(15) …Fred,            potatoes.
       NP                NP
     --------->T   ------------------------<T
     S/(S\NP)      (S\NP)\((S\NP)/NP)
     ----------------------------------------------*
```

After undergoing type-raising, *Fred* receives the designation *S/(S\NP)* – namely, that of a lexical item looking to its right for a verb seeking a subject to its left. In effect, *Fred* becomes a function over a one-argument verb. Similarly, *potatoes* receives the typing of an object looking to its left for a two-argument verb seeking an object to its right. However, combination of the two items is still blocked without the rule of forward mixing composition, which is shown below:

(16) Forward Mixing Composition ($>B_x$)

$[X/Y]_{\&} \quad Y\backslash Z \rightarrow_{\mathbf{B}} [X\backslash Z]_{\&}$

where Y = S\NP

Given this rule, the subject and object in the right conjunct may now combine to form a category of the type *S\((S\NP)/NP)*:

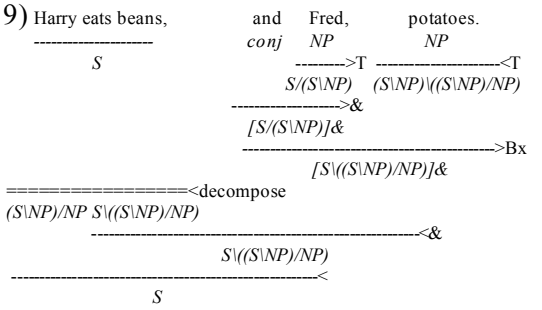
```
(17) Fred,            potatoes.
      NP                NP
    --------->T   ------------------------<T
    S/(S\NP)      (S\NP)\((S\NP)/NP)
    -------------------------------------->Bx
              S\((S\NP)/NP)
```

(Note: Marking the subject NP Fred for conjunction, which would have occurred prior to

(17), is not represented in the derivation.[1]*)*
Yet again, however, the derivation is blocked; with the category *S* on one side of the derivation – *Harry eats beans* – and the category *S\((S\NP)/NP)* on the other, there is no means by which the two conjuncts may combine (assuming the right conjunct is marked for coordination). In order to allow such combination, and in order to take into account semantic constraints on gapping, Steedman (1990) posits a rule of decomposition:

(18) Decomposition (<decompose)
X → Y X\Y
where X = S
and Y = given(X)

Decomposition requires that the category of the decomposed element be *S*, and that the other be provided in the discourse. This semantic discourse-sensitivity helps to limit the *Y* category in (18) to one that is relevant to a particular context. Without the decomposition rule, gapping sentences would be otherwise underivable in CCG; with the rule, we may finally derive the entire gapping sentence:

(19)


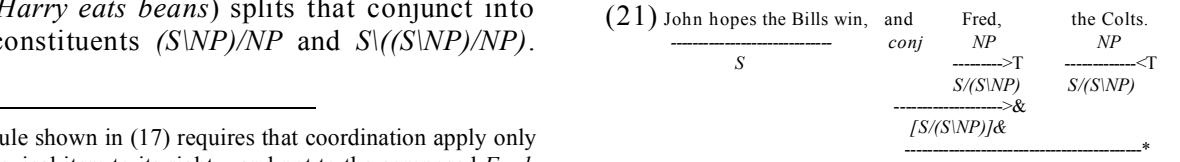
When decomposition of the left conjunct occurs, the verb is separated out from the subject and object (or, as I will show, the embedded subject). The decomposition rule applied to the left conjunct of a canonical gapping sentence (e.g. *Harry eats beans*) splits that conjunct into two constituents *(S\NP)/NP* and *S\((S\NP)/NP)*. The second of these constituents can be conjoined with the right conjunct since it is of identical type to produce another constituent of the same type. Finally, the result of coordination serves as the argument to the first of the decomposed constituents. The CCG analysis of gapping thus reflects the intuition that the verb (i.e., the *(S\NP)/NP* derived using decomposition) takes scope over both conjuncts.

4 Cross-clausal gapping

I will now consider data that pose a potential problem for most existing theories of gapping. Most of the data involves instances of what I term cross-clausal gapping, in which a gap ranges across an embedded clause, targeting the matrix and embedded verbs and leaving the subjects of both clauses as remnants. The relation in such gaps is thus one of subject/subject, rather than the typical subject/object relationship found in canonical instances of gapping. Two typical examples are produced below:

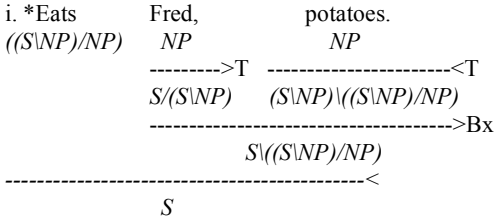
(20) a. John hopes the Bills win, and Fred, the Colts.
b. Robin knows a lot of reasons why dogs are good pets, and Leslie, cats. (C&J 2005:273)

We see, in both cases, that the sentence-final remnant is a subject – *the Colts* in (20a), and *cats* in (20b) – rather than an object, which causes the CCG derivation to break down. If the phrases in the second conjunct receive the typing of a traditional subject, namely, *NP* or, when type-raised, *S/(S\NP)*, we are left with no means of saving the derivation. If we type both subjects in the right conjunct with nominative Case (i.e. with the typical type-raised subject category), then the derivation fails as shown in (21):

(21)


The pair of string-adjacent subjects in the right conjunct cannot combine: even though the subjects have identical typing, and the subject *Fred* is marked for coordination, Steedman's coordination rule cannot save the derivation. As the forward coordination rule has already applied to mark *Fred* for coordination, the left-looking backward coordination rule must then apply; if both subjects in the right conjunct are typed with nominative Case, the derivation fails.

[1] The rule shown in (17) requires that coordination apply only to the lexical item to its right – and not to the composed *Fred, potatoes* constituent – in order to prevent the derivation of ungrammatical forms like the following, as noted by Steedman (1990):



However, CCG's machinery is fully capable of describing the data if we allow one crucial assumption: that the cross-clausal, sentence-final constituent in the second conjunct is typed as an object – just as CCG would predict for the sentence to be derived successfully.

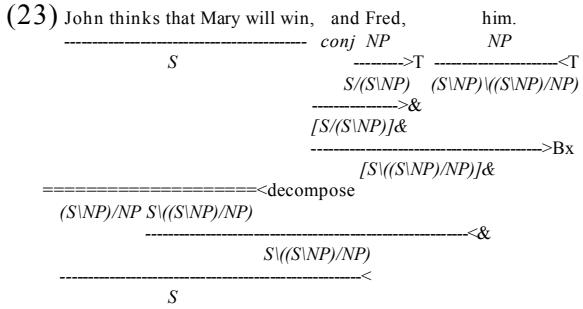
Though this assumption may initially seem somewhat *ad hoc*, evidence from Case lends support for this view. In English gapping sentences, there is a tendency to favor accusative pronominals in the second conjunct. Take the following data:

(22) a. John thinks (that) Mary will win, and Fred$_i$, him$_j$/*he$_j$/me/*I.[2]
b. I hope (that) Mary wins, and you, him/me.
c. John delivered a speech on why the Giants will win, and Fred, the Bills.

In (22a-b), the rightmost element in the right conjunct may only surface as a pronoun marked for accusative Case – the Case of a traditional object in English. To account for the possibility of subject extraction when *that* is not present in such examples, Steedman suggests that subjects can in some cases be analyzed as objects of the higher predicate. One might be tempted, therefore, to treat (22a) in this fashion, with the embedded subject in the left conjunct *Mary* analyzed as an object of *thinks*. There is reason to doubt this, however: In example (22c) the embedded subject cannot, in fact, be plausibly analyzed as the object of the higher verb. I follow Steedman in assuming that type raising is a reflection of case marking. When one type-raises a subject, for example, the resulting category is an *S/(S\NP)*, which effectively shifts the subject from being the argument of a verb phrase to instead being a function over a function over the argument of a verb phrase – that is, the subject becomes a function that takes a left-looking verb phrase as its own argument.

I posit that the typing of the sentence-final subject as an object points to a requirement that it bear accusative Case. The grammar permits the combination of a type-raised subject and object in CCG; we see quite clearly via empirical gapping data that such an allowance is necessary, both in English and in e.g. German (see Steedman 2000 for further discussion). A sample derivation, parallel to those of canonical English gapping sentences, is shown below:

(23) John thinks that Mary will win, and Fred, him.



The syntactic apparatus of CCG predicts that the rightmost element in an instance of cross-clausal subject/subject gapping must be typed as an object for the derivation to proceed. Thus, it should not be surprising for us to find the rightmost NP to surface with accusative Case – a surface representation of the fact that the rightmost element should receive object typing.

The account seems more plausible when one considers other, clearly ungrammatical cross-clausal gapping constructions. Oftentimes, the accusative-Case pronoun is required:

(24) a. John hopes (that) you win, and Fred, me/*I/him/her/*he/*she.
b. John delivered a speech on why Fred resigned, and Bill, me/I*/him/her/*he/she*.

In each of the examples above, the sentence-final remnant – interpreted in (24a) as the subject of the string ____ *win(s)*, and in (24b) as the subject of ____ *resigned* – may only appear in accusative Case.

One might argue that the appearance of accusative Case is simply due to the general unlikelihood of finding sentence-final subjects in English. Much past research (e.g. Schutze 2001) has commented on the status of accusative Case as default Case in English, and some would thereby conclude that this fact renders the analysis moot – that is, that the final subject defaults to accusative Case in the absence of an overt nominative Case assigner.[3] However, I contend that even if the default Case of English is accusative, or even if English favors accusative Case for sentence-final NPs, the analysis still follows: the tendency of English to

[2] Where *him* refers to another discourse-given individual. Example (22b) makes this relationship more apparent.

[3] An additional object of study would be the presence (or lack therof) of cross-clausal gapping in languages that necessarily mark nominative Case overtly, particularly in languages in which default Case is not accusative; if the analysis for English translates crosslinguistically, one would expect that the sentence-final subjects surface in accusative Case, both in instances of pronominals and in non-pronominal NPs.

favor accusative Case sentence-finally receives a surface manifestation via CCG typing.

5 Binding in cross-clausal gapping

Cross-clausal gapping sentences show an interesting property with respect to reflexive binding. Typically, a matrix subject cannot bind an embedded reflexive. However, cross-clausal gapping sentences like (25a-b) demonstrate the necessity of long-distance reflexive binding:

(25) a. *John$_i$ thinks that Mary is in love with himself$_i$.
b. John$_i$ thinks that Mary$_j$ is in love with Fred$_k$, and Bill$_l$, with himself$_{??i/??k/l}$/herself$_{?j}$/him$_{*l}$/her$_{*j}$.

As shown in (25a), in normal (non-gapping) circumstances, the matrix subject cannot bind the embedded reflexive. The most salient reading of sentence (25b) is the one in which Bill thinks that Mary loves him.[4] Under this cross-clausal reading, only the reflexive *himself*, referring to *Bill*, is completely acceptable. The use of *her* to refer to *Mary* is also ungrammatical. In the surface structure, there are no clause boundaries separating *Mary* from *her* – it is a Condition B violation. With CCG, there is no underlying syntactic structure or representation. The CCG notion of surface structure is essentially, 'what you see is what you get,' and the syntax builds up canonically non-standard constituents alongside a corresponding semantic interpretation. Even if the surface structure does not reflect the iteration of *Mary* that one would expect to be present underlyingly in an abstract account (i.e. in the second conjunct, as part of the deleted material), the presence of *Mary* in the first conjunct, together with the lack of a surface-level clause boundary, causes the non-reflexive *her* to be ungrammatical.

Thus, not only is long-distance reflexive binding available, it is in fact a necessity, as the unavailability of co-reference between *Bill* and *him* in (25b) indicates. Additional data provide further support for this picture:

(26) a. Fred thinks (that) Mary will win, and John, himself$_i$/??him$_i$/*he$_i$.
b. Fred believes (that) Mary loves John, and Bill$_i$, himself$_i$/*him$_i$.
c. Mary said (that) the stone had fallen on Sue, and Bill$_i$, on himself$_i$/*him$_i$.

Examples (26a-c) provide a different sort of evidence. As discussed previously, the rightmost element receives accusative case. We see that if this element is interpreted as coreferential with the leftmost element in the gapped constituent, the former must be a reflexive and not a pronoun. This is in contrast to what one would expect if the gapped material were reconstructed. Taking (26c) as an example, if binding occurs at the semantics, then some level of reconstruction of the gapped clause should be possible – and the reflexive should be dispreferred, because of the presence of a clause boundary intervening between *Fred* and *him/himself*. Yet, because the reflexive reading is not only available, but is in fact required, we receive evidence that the binding occurs at the level of the surface syntax, where *Fred* and *him/himself* are string-adjacent and fall within the same binding domain.

In practice, how might one represent binding in the syntax? Steedman (1996) treats binding at the level of predicate-argument structure. He assumes that bound anaphors receive a treatment that is syntactically identical to other NPs; the only difference is that they are marked with a *+ANA* feature. At the level of interpretation, which is built up simultaneously with the syntactic structure, reflexives are interpreted as a function of the type *self'* – the representation of *himself* is shown below:

(27) himself := $NP_{+ANA,3SM}$: *self'*

The Steedman approach is similar to the account of reflexives provided by Reinhart and Reuland (1993), in which the reflexive-marking is reflected on the verb. Given the above definition of the *-self* anaphor, the reflexivization of a transitive verb is thus represented with the following rule (from Steedman 1996):

(28) $(S\backslash NP_{agr})/NP : f \rightarrow (S\backslash NP_{agr})/NP_{+ANA,agr} : \lambda g.\lambda y.g\,f\,(ana'y)y$

In such a rule, the resulting predicate-argument structure is the function $gf(ana'y)y$, in which the variable g takes as its range the anaphoric interpretation of *self'*. It should be fairly plain to see how syntactic combination of verb and anaphor occurs derivationally – it proceeds as expected. In the interpretation structure, the semantic construction builds up parallel to its syntactic counterpart. This process restricts anaphoric binding to local domains, but it disallows the application of the λ-calculus to an already-composed constituent.

[4] Example (25b) requires some prosodic contrast (in which *John* and *Fred* receive emphasis, then *Bill* and *himself*) to make the cross-clausal reading completely clear, but the same is true for most gapping sentences.

Normally, this would not pose a problem – except in cases like those shown above, in which the cross-clausal gapping sentences demonstrate long-distance reflexive binding. Examples like (26a) require that binding occur at the level of the surface structure, which the Steedman account does not allow. Steedman's rule (28) could account for the cross-clausal cases, if it were able to apply following application of the decomposition rule (18) in the left conjunct; however, the rule (28) is strictly a lexical one and thus cannot apply following decomposition. Counterproposals exist that feature a different sort of reflexive binding – namely, one in which the reflexive is itself marked (rather than the verb) as a λ-operator that turns a two-argument function into a one-argument function, in effect reversing the normal function/argument structure. In such proposals, e.g. Szabolcsi (1989), the reflexive *W* is essentially a type-raised NP that causes an identity relation between arguments of a verb; this process is shown below (adapted from Szabolcsi 1989):

(29) $W = \lambda f.\lambda x.fxx$

(30) Assuming a transitive verb with interpretation $\lambda y.\lambda z.f\,(yz)$,
$\lambda f.\lambda x.[fxx](\lambda y.\lambda z.g\,[yz]) = \lambda x[gxx]$

This account differs crucially from the Steedman one in that it is the reflexive itself, and not the verb, on which reflexivization is marked. When building up cross-clausal gapping structures that contain long-distance reflexive binding, the Szabolcsi proposal would allow us to derive the proper binding facts for cross-clausal gapping structures; the result of combining the λ-terms for *John, himself* is a function that if given a transitive verb, will return a verb applied to both *John* and *himself* – precisely the intuitive reading of (26a).

Unfortunately, the Szabolcsi proposal runs into problems of its own. Given that it allows for long-distance reflexive binding, the Szabolcsi account overgenerates and permits the derivation of ungrammatical English sentences. If we type reflexives in the same way as other NPs in the syntax, then the proposal derives ungrammatical forms like the following:

(31) *John$_i$ thinks that Mary likes himself$_i$.

The Szabolcsi proposal allows reflexive binding to apply the λ-calculus to composed constituents; without restrictions on the application of binding to composed constituents, examples like (31) are predicted to be grammatical. Thus, although a proposal like the Szabolcsi one is necessary to account for the facts of cross-clausal gapping, it predicts that other long-distance binding will be grammatical, as well. As such, although the Szabolcsi theory is better able to capture the cross-clausal facts than the Steedman one, it fails to earn an unqualified endorsement as a preferred proposal overall.

6 Chinese type-raising

The previous discussion has centered almost exclusively on gapping in English. In the following section, the discussion shifts to a range of data in Chinese that can bring something to bear on the present analysis of gapping.

Wu (2002) adduces a class of gapping-like constructions in Chinese that display interesting behavior. Specifically, Wu shows that instances of gapping in Chinese are restricted to NP objects that carry some form of quantificational force, generally as part of a classifier phrase (a phenomenon also discussed in Li 1988 and Paul 1999). The alternation in grammaticality between quantified NP objects and bare nouns is shown below:

(32) a. Zhangsan chi-le san-ge pingguo, Lisi ~~chi-le~~ si-ge juzi.
Zhangsan ate three-CL apple Lisi ~~ate~~ four-CL orange
'Zhangsan ate three apples and Lisi four oranges.'
(Li 1988:41)

b. Zhangsan xihuan pingguo, Lisi *(xihuan) juzi.
Zhangsan like apple Lisi like orange
'Zhangsan likes apples and Lisi oranges.' (Wu 2002:3)

I take this restriction to demonstrate that type-raising in Chinese is restricted to a specific set of words and phrases, i.e. those carrying quantificational force. Throughout the development of CCG, type-raising has been accepted to occur relatively freely; however, I contend that the alternations in grammaticality in Chinese, and the impermissibility of gapping in sentences where the object lacks quantificational force, support the conclusion that type-raising in Chinese is in fact restricted. Given such a restriction, the facts of gapping in Chinese fall out naturally. Looking to (32a), we see that the object *juzi* "orange" is preceded by *si-ge* "four," also marked as a classifier phrase. The alternation in grammaticality between (32a), with a quantified object, and (32b), which contains the bare NP object *juzi*, is striking, and it provides a minimal pair for the analysis. Simply, when the object is not preceded by a quantificational element – in this case, a classifier phrase – gapping is unavailable.

Representing this fact in CCG is simple. Lexical items are assigned specific categories, and the derivation of gapping sentences requires a highly specific category assignment that allows for the combination of subject and object in the right conjunct via forward crossing composition. In order for the subject and object to combine, each item must be type-shifted from an argument to a function-over-functions-over-arguments. A parametric constraint on the type-shifting of bare NP objects means that in examples like (32b), the subject and object cannot combine, and the derivation crashes. This derivational crash is shown below[5]:

(33) *Zhangsan xihuan pingguo, Lisi juzi.
Zhangsan like apple Lisi orange
---------- *(conj)* *NP* *NP*
S ---------->T --------*
S/(S\NP)

The method for deriving the licit Chinese example (32a) should by now be equally apparent (shown on the next page):

(34) Zhangsan chi-le san-ge pingguo, Lisi ~~chi-le~~ si-ge juzi.
Zhangsan ate three-CL apple Lisi ~~ate~~ four-CL orange
---------- *(conj)* *NP* *NP*
S -------->T ----------<T
S/(S\NP) *(S\NP)\((S\NP)/NP)*
---------->&
[S/(S\NP)]&
---------->Bx
==========<decompose *[S\((S\NP)/NP)]&*
(S\NP)/NP *S\((S\NP)/NP)*
----------<&
S\((S\NP)/NP)
----------<
S

Due to the presence of the classifier phrase, the NP object *juzi* may type-shift. It combines with the subject *Lisi* to produce a subject-object constituent, which subsequently combines with the result of the decomposition in the left conjunct. The decomposition separates out the verb from the subject-object constituent, which is then coordinated with the analogous subject-object constituent in the right conjunct. Finally, the coordinated subject-object constituent combines with the transitive verb through function application, and the result is a well-formed sentence. Gapping in Chinese thus depends on the expression of the NP object and its (in)ability to type-shift.

The notion of crosslinguistic variation in the range of NP interpretations is not a new one; Chierchia (1998) established a system for defining languages in terms of the availability of mass and count nouns, and the expression of each. Chierchia described two features – ±argument and ±predicate – to define the expression of nouns. In some instances, all nouns are arguments (meaning that bare nouns occur freely, as in Chinese); in others, all nouns are predicates (bare nouns are practically, if not totally, excluded, as in French); finally, in still other instances nouns may be either predicates or arguments. This final category (which includes English) allows for greater freedom in type-shifting of phrasal projections. Chierchia applies this expression of type-shifting specifically to mass/count noun distinctions, but the same principle informs our conception of type-shifting in CCG – some languages, like English, do allow type-shifting to occur freely. In others, e.g. Chinese, type-shifting is restricted; simply, bare NPs cannot type-shift.

Chinese allows type-raising only in case there is some quantificational force inherent in the DP; with this single observation, we see that the facts of gapping in Chinese, and the environments in which it is permissible, follow as a natural consequence of the CCG theory.

7 Conclusion

In this paper, I have outlined a number of facts that any theory of gapping must analyze. The CCG proposal of Steedman (1990, 2000), carries a high degree of predictive power in managing the wide variety of cross-clausal gapping data contained herein. CCG predicts the typing of the rightmost subject in cross-clausal gapping data as an object; evidence from Case supports this hypothesis. Reflexive binding in cross-clausal structures favors the Szabolcsi (1989) proposal, in which binding occurs at the level of the surface structure. Additionally, facts from Chinese buttress the CCG analysis, as its NP category-assignment delivers a straightforward explanation for the ungrammaticality of gapping sentences containing non-quantified NP objects: simply, they are unable to undergo type-shifting.

Acknowledgments

Many thanks are due to Robert Frank, Raffaella Zanuttini, fellow linguists in LING 491, friends, family and many others.

[5] I assume the presence of coordination on some (perhaps syntactic) level; similar examples, in which a coordinator fails to surface overtly, are also found in English (see example (3)). Such examples have long been noted (see Sag 1976), and I see them as posing no real threat to the CCG analysis of gapping.

References

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6:339-405.

Culicover, Peter W., and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.

Li, M.D. (1988) *Anaphoric Structures of Chinese*. Taipei: Student Book Co, Ltd.

Paul, W. 1999. Verb gapping in Chinese: a case of verb raising. *Lingua* 107:207-226.

Reinhart, Tanya, and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24:657–720.

Ross, John. 1967. Constraints on variables in syntax. Doctoral Dissertation. Massachusetts Institute of Technology.

Ross, John. 1970. Gapping and the order of constituents. In Bierwisch and Heidolph (eds.), *Progress in linguistics*. The Hague: Mouton. 249-259.

Sag, Ivan. 1976. Deletion and logical form. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Schütze, Carson. 2001. On the nature of default Case. *Syntax* 4:205-238.

Steedman, Mark. 1990. Gapping as constituent coordination. *Linguistics and Philosophy* 13. 207–263.

Steedman, Mark. 1996. *Surface structure and interpretation*. Cambridge, Mass.: MIT Press.

Steedman, Mark. 2000. *The syntactic process*. Cambridge, MA: MIT Press.

Szabolcsi, A. 1989. Bound variables in syntax (are there any?). In Bartsch, van Benthem, and van Emde Boas (eds.), *Semantics and contextual expression*. Dordrecht: Foris. 295-318.

Wu, Hsiao-hung Iris. 2002. On ellipsis and gapping in Mandarin Chinese. Master thesis. National Tsing Hua University.