# Chapter 10

## Symmetry and Asymmetry in Left and Right Extraction

The possibility of conjunction offers one of the best criteria for the initial determination of phrase structure.
—*Syntactic Structure* Noam Chomsky, 1957:36

The reduction of both *wh*-extraction and right-node–raising of an element $X$ to contiguous combination or merger depends in general on the possibility of making the residue of both into a unbounded constituent of type $S|X$ by identical processes of function composition, with or without coordination of like types. It follows that the present theory makes a broad prediction of symmetry for left- and right-extraction in English: whatever can undergo *wh*-extraction from the periphery of a typable constituent can potentially right-node–raise from it, and vice versa. subject to the same constraints. Conversely, if an element cannot right node raise, then it should not be able to left extract. A number of instances of this tendency were noted above, including that of extraction from English embedded *that*-complements, where embedded objects can both *wh*-extract and right-node raise, with identical residues of type $S/NP$, but where subjects of sentential complements can in general extract neither to the left nor to the right.[1]

However, there is a difference. English leftward-moving categories are special-purpose non-order-preserving exotypic categories like those in (18) of chapter 9, distinct from the order-preserving type-raised cased rightward movers, permitting exceptions to symmetry, all of which must under present assumptions be specified in the English lexicon, and as such be expected to vary across other languages.

For example, we also saw in section 9.6 that English subjects of bare complements lacking a complementizer can wh-extract, but where there is no corresponding in situ argument category that would permit right-node raising or Heavy NP-shift:

(1) a. Who did they say won the election?
b. *They said won the election the heaviest candidate.

The following are some further cases where symmetry might be expected, but where

1. The use of a single mechanism in leftward and rightward extraction was a feature of the first version of GPSG (Gazdar, 1981). This very attractive feature of the theory was almost immediately abandoned for theory-internal reasons that do not apply to the present theory under the Projective Dependency Principle (PDP) (1) of chapter 9.

exceptions to symmetry have been claimed.

## 10.1 Across-the-Board Extraction

Non-traditional unbounded constituents of the type $S/NP$ that have most recently been encountered at length in chapter 9.4 as the type of the residue of relativization also showed up in an earlier chapter as the residue of rightward extraction, or right-node raising, as in (2a), whose derivation is shown in figure 10.1.

(2) Freeman admires, and Hardy says Willis detests, sincerity/#it.

There are two further points to note about the above example and the derivation in figure 10.1. First, the derivation crucially depends on the availability of composition merge and type-raising. Second, such right-node-raised objects must be grapho-phonologically "heavy" and "rhematic" or discourse-new.

For example, the pronoun alternative in (2), which is phonologically "light" and by definition discourse-given, is unacceptable in this construction. We can capture this fact by giving English pronouns only the "clitic" accusative category $(S\backslash NP)\backslash((S\backslash NP)/NP)$ and excluding the right-node-raising and Heavy-NP-Shifting category $S\backslash(S/NP)$ for them. Thus pronouns can only take part in traditional transitive derivations like (22) of chapter 4.1, and not in the nonstandard derivation (23) there.[2]

As in the GPSG account of Gazdar 1981, ã type-dependent account of extraction and coordination, as opposed to standard accounts using structure-dependent rules, makes the across-the-board condition (ATB) on extractions from coordinate structure a prediction for both lefteard and rightward extraction, without any additional stipulation of parallel structure conditions (Goodall, 1987) or syntactic multidominance (McCawley, 1982, 1987; Citko, 2017), as the following examples reveal:

(3) a. A saxophonist [that$_{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)}$ [[Harry admires]$_{S/NP}$ and [Louise says she detests]$_{S/NP}$]$_{S/NP}$]$_{N\backslash_{\diamond\star}N}$

b. A saxophonist that$_{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)}$ *[[Harry admires]$_{S/NP}$ and [Louise says she detests him]$_{S}$]]

c. A saxophonist that$_{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)}$ *[[Harry admires him]$_{S}$ and [Louise says she detests]$_{S/NP}$]

2. This claim is borne out by the fact that pronouns generally fail to attract nuclear pitch-accents in the spoken language and cannot take part in the prosodic structure indicated in (2c), as predicted by the analysis of intonation structure in Steedman 2000a, 2014 and chapter 6.

$$
\dfrac{\dfrac{\dfrac{[\text{Freeman admires}]}{\begin{array}{c}S/NP\\ :\lambda x.admires\,x\,freeman\end{array}}{>}\mathbf{B} \quad \dfrac{\text{and}}{\begin{array}{c}(\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T}\\ :\lambda f\lambda g\lambda x.fx\sqcap gx\end{array}} \quad \dfrac{[\text{Hardy says Willis detests}]}{\begin{array}{c}S/NP\\ :\lambda x.says(detests\,x\,willis)hardy\end{array}}{>}\mathbf{B}}{\begin{array}{c}(S/NP)\backslash_{\star}(S/NP)\\ :\lambda g.\lambda x.g\,x\wedge says(detests\,x\,willis)hardy\end{array}}{>}}{\begin{array}{c}(S/NP):\lambda x.admires\,x\,freeman'\wedge says(detests\,x\,willis)hardy\end{array}}{>} \quad \dfrac{\text{sincerity}}{\begin{array}{c}S\backslash(S/NP)\\ :\lambda p.p\,sincerity'\end{array}}
$$

$$
\begin{array}{c}S:admires\,sincerity\,freeman'\wedge says(detests\,sincerity\,willis)hardy\end{array}{<}
$$

Figure 10.1: Unbounded right-node raising

(4) [Harry admires $_{S/NP}$ and [Louise says she detests]$_{S/NP}$]$_{S/NP}$ [the saxophonist]$_{S\backslash(S/NP)}$.

a. *[[Harry admires]$_{S/NP}$ and [Louise says she detests him]$_S$] [the saxophonist]$_{S\backslash(S/NP)}$.

b. A saxophonist that$_{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)}$ *[[Harry admires him]$_S$ and [Louise says she detests]$_{S/NP}$] [the saxophonist]$_{S\backslash(S/NP)}$.

The theory also predicts the ill-formedness of the following violation of the "same-case condition" on the ATB exception to the Coordinate Structure Constraint, since the right conjunct is not of the same CCG type as the left conjunct:

(5) a. *A saxophonist that$_{(N\backslash_{\diamond\star}N)/_{\diamond\star}(S/NP)}$ [[Harry admires]$_{S/NP}$ and [detests bossa nova]$_{S\backslash NP}$]$_*$

b. *[[Harry admires]$_{S/NP}$ and [detests bossa nova]$_{S\backslash NP}$]$_*$ [the man in the Brooks Brothers shirt]$_*$

However, in the case of (6a) there is another derivation, in which *Harry admires* is analyzed as a bare relative under the mechanism set out in section 9.5, analogous to the unreduced relative in (6b)

(6) a. ?A saxophonist [[that detests bossa nova]$_{N\backslash N}$ and [Harry admires]$_{N\backslash N}$ ]$_{N\backslash N}$

b. ?A saxophonist [[that detests bossa nova]$_{N\backslash N}$ and [that Harry admires]$_{N\backslash N}$ ]$_{N\backslash N}$

c. *[[Detests bossa nova$_{S\backslash NP}$ and [Harry admires]$_{S/NP}$]$_*$ [the man in the Brooks Brothers shirt]$_*$

Since the fragment *Harry admires* can and must be analyzed as bare relative $N\backslash N$, for some speakers (6a) escapes the same-case condition via the "back-door" of this alternative derivation. This analysis is not available for the right-node raised version (6c).

Finally, the fact that *embedded* subject extraction from bare complements of verbs like *think* can happily coordinate with object extraction confirms the analysis of embedded subject extraction in section (9.6) as arising from *rightward* subcategorization for NP$_{+wh}$ of the matrix verb:

(7) a. A saxophonist that some critic [thought was a genius]$_{(S\backslash NP)/NP_{+wh}}$ and [praised]$_{(S\backslash NP)/NP}$.

b. A saxophonist that some critic [praised]$_{(S\backslash NP)/NP}$ and [thought was a genius]$_{(S\backslash NP)/NP_{+wh}}$.

Despite the overwhelming evidence that coordination is a rule operating over like types in the strictest sense, it has sometimes been suggested that the Coordinate Structure Constraint and the ATB exception are an illusion. The suggestion is most influ-

entially made on the basis of some examples first noticed by Ross (1967), Goldsmith (1985), and Lakoff (1986) like the following:

(8) a. What did you [go]$_{VP}$ and [see]$_{VP/NP}$?
b. This is the stuff that people in the Caucasus [drink every day]$_{VP/NP}$ and [live to be a hundred]]$_{VP}$.

Such examples have been used by Kehler (2002) Asudeh and Crouch (2002) and (Zhang, 2010, 135-139) to argue against the reality of the ATB generalization.

Ross and Goldsmith themselves argued that extractions like (8a) involve a lexicalized multiword aspectual item "go and", which in present terms must bear the category $VP/VP$ (and tensed etc. variants). They also argued that cases like (8b) involve another, noncoordinate, subordinating lexical category for *and*, and as such do not constitute counterexamples to the CSC and ATB constraints. They note the presuppositional and volitional semantics of the sentences in question (and the absence of such overtones from true coordinates), as well as the fact that—as Postal 1998 points out—no *other* conjunctions support such extractions. Thus:

(9) a. What did you go to see?
b. This is the stuff that people in the Caucasus drink every day so that they live to be a hundred.

(10) a. *What did you go or see?
b. *This is the stuff that those guys in the Caucasus drink every day or else die young.

Nor are such leftward extractions mirrored by equivalent right-node raising, as in (11a), unlike the standrd across-the-board cases like (11b):

(11) a. *Those guys in the Caucasus drink every day and live to be a hundred a kind of fermented mare's milk.
b. Harry admires and Louise says she detests several very famous saxophonists.

The suggestion by these authors of additional asymmetric categories for "and" was tentatively adopted in *TS*was tentatively adopted in *TS*. However, the latter observation makes these sentences look akin to the phenomenon of so-called "subject gapping", seen in English examples like the following, in which the right conjunct is not adjacent to a left conjunct of the same type:

(12) Through the door bounded a huge dog and [bit me in the leg]$_{S\backslash MP}$.

The phenomenon is rather marginal in English, but it is very common in German and Dutch. Heycock and Kroch (1994) note the relation of subject-gapping to the apparent violation of the ATB constraint. Subject gapping is analysed in Steedman, 2017 as a form of stripping coordination.

## 10.2 Right-edge restrictions

The following related asymmetry has led Wilder, 1999:(5a) to postulate a Right Edge Restriction (RER) on right-node–raising from which ATB *wh*-extraction is exempt (ibid.:(5b)):

(13) a. *I gave a present and congratulated all the winners.
b. The man who I gave a present and congratulated.

Related examples were the reason for GPSG abandoning Gazdar's 1981 claim that RNR was mediated by the same mechanism as *wh*-extraction, and to HPSG embracing a deletion/ellipsis analysis (Beavers and Sag, 2004). However, we saw for (8) and (9) that "gave a present" is not a constituent with the type of a transitive verb, as evidenced by the fact that, unlike all other double-complement verbs (with the possible exception of "promise" Bach, 1979, 1980), Heavy NP shift, which involves the same composition as medial RNR) is blocked for the ditransitive:

(14) a. *I gave a present all the winners.
b. I persuaded to take a bath all the winners

Instead, (13b) can be allowed, and (13a) excluded, by an antecedent-governed lexical category for ditransitive verbs, parallel to the embedded subject-extracting bare complement verb (**??**):[3]

(15) gave:= $((S\backslash NP)/NP_{ANT})/NP) : \lambda w \lambda x \lambda y.gave\, w\, x\, y$

The derivation of (13b) is then as follows:

(16)

| man | who | I | gave | a present | and | congratulated |
|---|---|---|---|---|---|---|
| $N$ | $(N\backslash N)/(S/NP_{ant})$ | $S/(S\backslash NP)$ | $((S\backslash NP)/NP_{ANT})/NP$ | $NP^{\uparrow}$ | $(X\backslash_{\star}X)/_{\star}X$ | $(S\backslash NP)/NP$ |

$$((S\backslash NP)/NP_{ANT})/NP \quad NP^{\uparrow} \;\; <$$
$$(S\backslash NP)/NP_{ANT}) \quad (X\backslash_{\star}X)/_{\star}X \quad (S\backslash NP)/NP \;\; <>$$
$$(S\backslash NP)/NP_{ANT} \quad (S\backslash NP)/NP_{ANT} \;\; >\mathbf{B}$$
$$S/NP_{ANT} \quad (S\backslash NP)/NP_{ANT} \;\; >$$
$$N\backslash N \quad S/NP_{ANT} \;\; >$$
$$N \quad N\backslash N \;\; <$$

(13a) is disallowed, because the right-node-raised NP is lexical, rather than *ANT*-marked.

The present analysis has the advantage that we do not need to stipulate exemptions from RER for (13b), and examples like the following, which RER would otherwise appear to exclude:

3. As one would expect, there are dialects of English that lack this special-case category and simply disallow (13b).

(17) a. I persuaded to take a bow and congratulated all the winners.
b. I sold to the library and Mary donated to the museum several very valuable books

Wilder's related example (18) is blocked under present assumptions, because the type of "Joss will donate to the library" and of "Mary will donate" have to be the same (that is, either $S/NP$ or $(S/PP)/NP$) for coordination to apply and RNR to be possible.

(18) *[Joss will donate $\__i$ to the library and Mary will donate $\__i]_{S/NP}$ several old novels$_i]_S$] to the museum

In the former case, "to the museum" would have to distribute over "John will donate to the library, and Mary will donate, several old novels" (cf. Citko and Gračanin-Yuksek, 2020:36):

Wilder analyses (19a) as object RNR, conflicting with the present account:

(19) a. John should [fetch]$_i$ and [give the book$_i$ to Mary].
b. John should [fetch]$_{i,j}$ and [give]$_{i,j}$ the book$_i$ to Mary$_j$.

However, the analysis (b), in which "the book" and "to Mary" are right-node-raised out of both conjuncts is also available, and seems to correspond to the only available reading.[4]

## 10.3 Asymmetric islands

: There have similarly been many claims in the literature since Wexler and Culicover (1980:299) that left- and right- movement are not symmetrical with respect to island constraints (Beavers and Sag, 2004; Cann, Kempson, and Marten, 2005; Sabbagh, 2007; Citko and Gračanin-Yuksek, 2020. See Postal, 1998, Steedman, 2012:101-103, Bachrach and Katzir, 2009, and Kubota and Levine, 2020:325-327 for counter-arguments.)

Examples like the following are not entirely compelling (the judgments are Beavers and Sag's), especially when care is taken to make intonational prosody the same in (a) as in (b) (Steedman, 2000a, 2012:103):

(20) a. ??Those unflattering pictures of Qaddafi, Yo knows several men who buy, and Jan knows several men who sell.
b. Yo knows several men who buy, and Jan knows several men who sell, those unflattering pictures of Qaddafi.

It seems possible that the asymmetries between *wh*- and RNR–extraction claimed (with a degree of uncertainty) by Citko and Gračanin-Yuksek (2020:ch.3) for Slavic

4. A similar analysis appears to apply to related examples discussed by Whitman (2009). See Kubota and Levine, 2020 for discussion and an alternative categorial analysis of coordination to the present one.

languages are similarly discourse-sensitive, rather than reflecting any fundamental difference in the nature of the long-range dependencies involved. Sabbagh, citing an anonymous reviewer, offers the following apparent violation of the adjunct island constraint:

(21) Politicians win when they defend, and lose when they attack, the right of a woman to an abortion.

Again, the judgment is his, but the example seems no better and no worse than the corresponding left-extraction "What right do politicians win when they defend and lose when they attack?". The same seems to be true of the "non-coordinate RNR" examples discussed by Hudson (1976); Postal (1994), and Phillips (2003). Bachrach and Katzir (2009) and Hirsch and Wagner (2015) discuss related examples of ATB *wh*-extraction out of islands. The lack noted earlier of a clear distinction between strong and weak islands, and the dependence of the latter on pragmatic factors, make it hard to draw any conclusion concerning asymmetry in left/right extraction from these data.

Examples like (22a) require "focal" stress or accent on the stranded prepositions, as indicated by capitals, and that similar intonated conjoined fragments also seem to license wh-extraction, as in (22b) (Ambridge and Goldberg, 2008; Chaves, 2012):

(22) a. CHESTER likes the person who visited us FROM, and LESTER likes the person who gave us a ticket TO, the beautiful island of Capri.
b. A place that CHESTER likes the person who visited us FROM, and LESTER likes the person who gave us a ticket TO.

Beavers and Sag, 2004 also note, following Davis (1992), that in Hausa, an SVO language with object pro-drop, while ATB object *wh*-extraction is allowed, RNR is not (Davis, 15, 16). However, the availability of object pro-drop in Hausa means that phonological emptiness cannot be taken as evidence of movement per se. If Hausa "movement" is really left-dislocation with pro-drop then there may be asymmetries with respect to discourse characteristics of left- and right-dislocated elements, with the former being by definition discourse-old as required by pro, but the latter required to be new or contrastive, as in English RNR. (This suggestion seems consistent with Davis's own analysis of finite and non-finite verbs in Hausa (ibid:(22)).) Related considerations may explain the asymmetry noted by McCloskey (1986) for Irish prepositions, which engender obligatory pro-drop, and "strand" for right-node raising, but not for *wh*-extraction (see below, although *pro* can act as resumptive Legate 1999:(11)).

## 10.4 Ross's Argument/Adjunct Cluster Coordination Asymmetry

One of the clearest confirmations that type raising is a universal feature of natural language morphosyntax related to case comes from the phenomenon of so-called nonconstituent coordination:

(23) I showed Chester Harry, and Lester Louise.

The derivation is written in full including the semantics as in figure 10.2a, from which it will be evident that the argument clusters (and type-raised categories in general) bear logical forms reminiscent of the "VP-shell" constituents postulated for the ditransitive by Larson (1988, 1990). The construction was noted by Williams (1981) as posing serious problems for the related but differently type-driven GPSG account of coordination.

The semantics of type raising, composition, and coordination given earlier at (11) of chapter 2, (6) of chapter 9.4, and (2) of this chapter guarantees that the derivation delivers the same logical form as *I showed Chester Harry and I showed Lester Louise*, as shown. The analysis also has the important property of allowing a certain amount of attested nonparallelism. For example, the number of adjuncts included in coordinating clusters may differ, as in the following real-life example from the Penn Wall Street Journal treebank, in which the first conjunct includes one adjunct to the verb *exceed*, and the second, two:

(24) Inflation will exceed [10% if the Fed doesn't take action], and [3% for a time even if it does].

The derivation is parallel to (20) and figure 10.2a, apart from involvement of adjuncts rather than the raised object. The crucial part of the derivation is shown in figure 10.2b.

The fact that the finite and infinitival complements can also undergo cluster coordination in examples like the following shows that they too have categories raised over the matrix verbs, like other argument types such as NP:

(25) a. I told Harry that it was raining and Sally that it would snow.
b. I told Sally to take a bath and Harry to have a wash.
c. I saw Harry come and Sally go.

However, the fact that extraction is possible out of finite and infinitival complements means that they also retain the unraised type, unlike other argument types like NP:

(26) a. The house that I told Harry that I would buy.
b. The bath that I told Sally to take.
c. The horse that I saw Harry steal.
d. *The horse that I met a man who stole.

Ross (1970) pointed out that this construction illustrates a universal asymmetry across languages, namely that rightward arguments and enclitic adjunct clusters form as right conjuncts, while leftward arguments and proclitic adjuncts form as left conjuncts. For example, in Welsh, a VSO language, clusters are also right conjuncts, as in figure 10.2b. Borsley et al., 2007:52 note that this derivation consitutes a movement-free version of an across-the-board head movement or verb-fronting account of the construction in a theory like that of Roberts (2005) deriving Welsh VSO surface orders

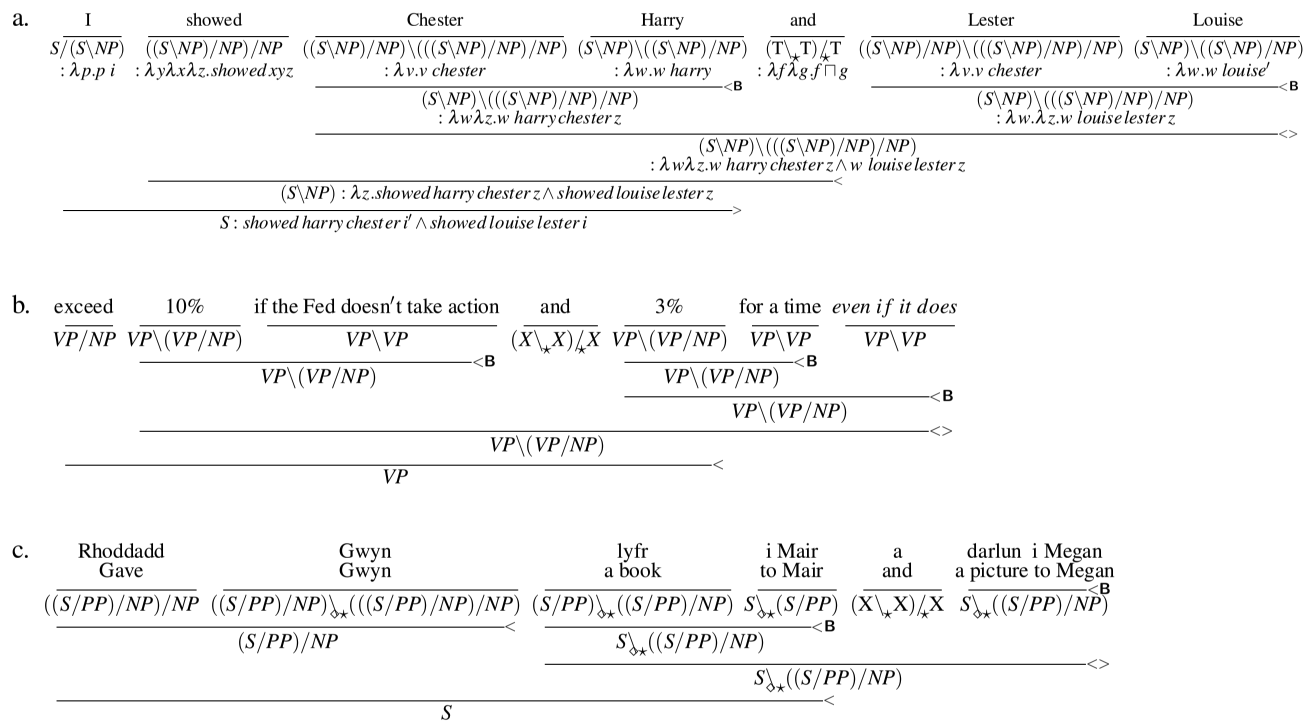
a.
I
showed
Chester
Harry
and
Lester
Louise
b.
exceed
10%
if the Fed doesn't take action
and
3%
for a time
even if it does
c.
Rhoddadd
Gave
Gwyn
Gwyn
lyfr
a book
i Mair
to Mair
a
and
darlun i Megan
a picture to Megan

Figure 10.2:

from underlying SVO, and is akin to the remnant-movement examples discussed in section 9.9.

Japanese, being a verb-final language, provides an example of leftward cluster-coordination, in which coordination itself requires a syntagmatic combinatory rule:[5]

(27)

| Ken – ga | Naomi – o, | Erika – ga | Sara – o | tazuneta |
|---|---|---|---|---|
| Ken.NOM | Naomi.ACC, | Erika.NOM | Sara.ACC | visited |
| $S/(S\backslash NP_{nom})$ | $(S\backslash NP_{nom})/((S\backslash NP_{nom})/NP_{acc})$ | $S/(S\backslash NP_{nom})$ | $(S\backslash NP_{nom})/((S\backslash NP_{nom})/NP_{acc})$ | $(S\backslash NP_{nom})/NP_{acc}$ |
| | $>\mathbf{B}$ | | $>\mathbf{B}$ | |
| $S/((S\backslash NP_{nom})/NP_{acc})$ | | $S/((S\backslash NP_{nom})/NP_{acc})$ | | |
| | | $\&$ | | |
| $S/((S\backslash NP_{nom})/NP_{acc})$ | | | | |
| | $>$ | | | |
| $S$ | | | | |

"Ken visited Naomi, and Erika Sara"

With more arguments, there are further derivational possibilities:

(28) Kyooju-ga gakusee-ni komonjo-o
[⟨Professor-NOM {student-DAT⟩ manuscript-ACC}]
miseta
showed$_{((S\backslash NP_{NOM})\backslash NP_{DAT})\backslash NP_{ACC}}$
'The professor showed the student the manuscript.'

The argument cluster *Kyooju-ga gakusee-ni komonjo-o* can form a constituent of type $S/(((S\backslash NP_{NOM})\backslash NP_{DAT})\backslash NP_{ACC})$ (indicated by [] brackets), which can coordinate with another preverbal argument cluster of the same type before applying to the ditransitive verb *miseta* on the right.[6]

Since there are three arguments in the cluster, there are two possible derivations of the cluster (respectively indicated by ⟨⟩ and {} brackets around binary clusters). The binary subunits themselves can coordinate within the cluster as a whole. All these variants are semantically correct, for the same reason figure 10.2 is. The implications for incremental parsing and interpretation of verb-final languages are discussed in appendix C.

The fact that the cross-linguistic availability of argument/adjunct-cluster coordination is a prediction of CCG is one of its major explanatory achievements. The phenomenon was among those leading Pesetsky (1995:175) to postulate "cascade struc-

5. Scrambled argument orders like (ia) can apparently also also undergo cluster coordination in Japanese:
(i) a. Naomi-o Ken-ga, Sara-o Erika-ga tazuneta.
b. ?Naomi-o Ken-ga, Erika-ga Sara-o azuneta.
This observation might seem to call into question the account of scrambling in terms of crossed composition in chapter 5, and to suggest that Japanese verbs are simply lexically ambiguous as to argument order, as wa suggested by Steedman and Baldridge (2011). However, clusters that would under that assumption have different types can also coordinate, as in (ib). This suggests that something more is going on in these examples, as indeed is the case for the English gapped sentence offered as a translation in (27).

6. Whitelock (1991) discusses prosodic evidence for such argument-cluster constituents and a corresponding left-branching derivational analysis of the Japanese clause.

ture as an additional level of constituent structure in addition to standard "layered" syntactic structure. Under the present systems these are alternatives at a single non-representational level of derivation structure.

The fact that the inclusion of order-preserving type raising further predicts Ross's generalization and the non-existence of the unattested class of languages with the same verbal lexical types as Japanese, but allowing the English pattern of cluster coordination, or with the same lexical types as English, but allowing Japanese-style leftward cluster coordination like the following, is similarly important (see Steedman 1985; Dowty 1985/1988).

(29) *Chester Harry and I showed Lester Louise.

## 10.5 Preposition Stranding and Coordination

The stranding prepositional category (65) of chapter 9.4 yields derivations such as those in figure 10.3 under coordination:[7]

The related examples in figure 10.4 are problematic for any account that (explicitly or implicitly) treats the category of English prepositions as $PP/NP$ or $PP^{\uparrow}/NP$—that is as specifier of argument PP.

In English, we have seen that proposition-stranding is possible under both leftward and rightward extraction. However, in reaction to the account of extraction in Gazdar (1981), which predicts a similar symmetry, McCloskey (1986) pointed out that Modern Irish, which never allows preposition stranding under relativization, does nevertheless allow it under Right Node Raising (McCloskey (1)):[8]

(30)

| | | | | | | |
|---|---|---|---|---|---|---|
| Níl | sé | in aghaidh | an | dlí | a thuilleadh | a bheith |
| is-not | it | against | the | law | anymore | be.INF |
| [ag éisteacht le | | nó | ag breathnú ar] | | | |
| [listen.PROG with | | or | look.PROG on] | | | |
| ráidió | agus | teilifís | an | Iarthair. | | |
| radio | and | television | the | West.GEN | | |

"It is no longer against the law [to listen to or look at] Western radio and television"

This possibility can be captured by assuming that Irish does have the English

7. Right-node-raising of the kind exhibited in figure 10.3c is pragmatically demanding, and has consequences for the intonational prosody in which such sentences are uttered. It is a virtue in the theory that the nonstandard constituency that supports such analyses of coordinate structure, define exactly the notion of constituency that is required to support those intonation structures, as discussed in chapter 6.

8. McCloskey claims that preposition stranding is not possible with Heavy NP shift in Irish. However, we have seen in section 9.8 that the same holds for English preposition stranding ander the preent analysis.

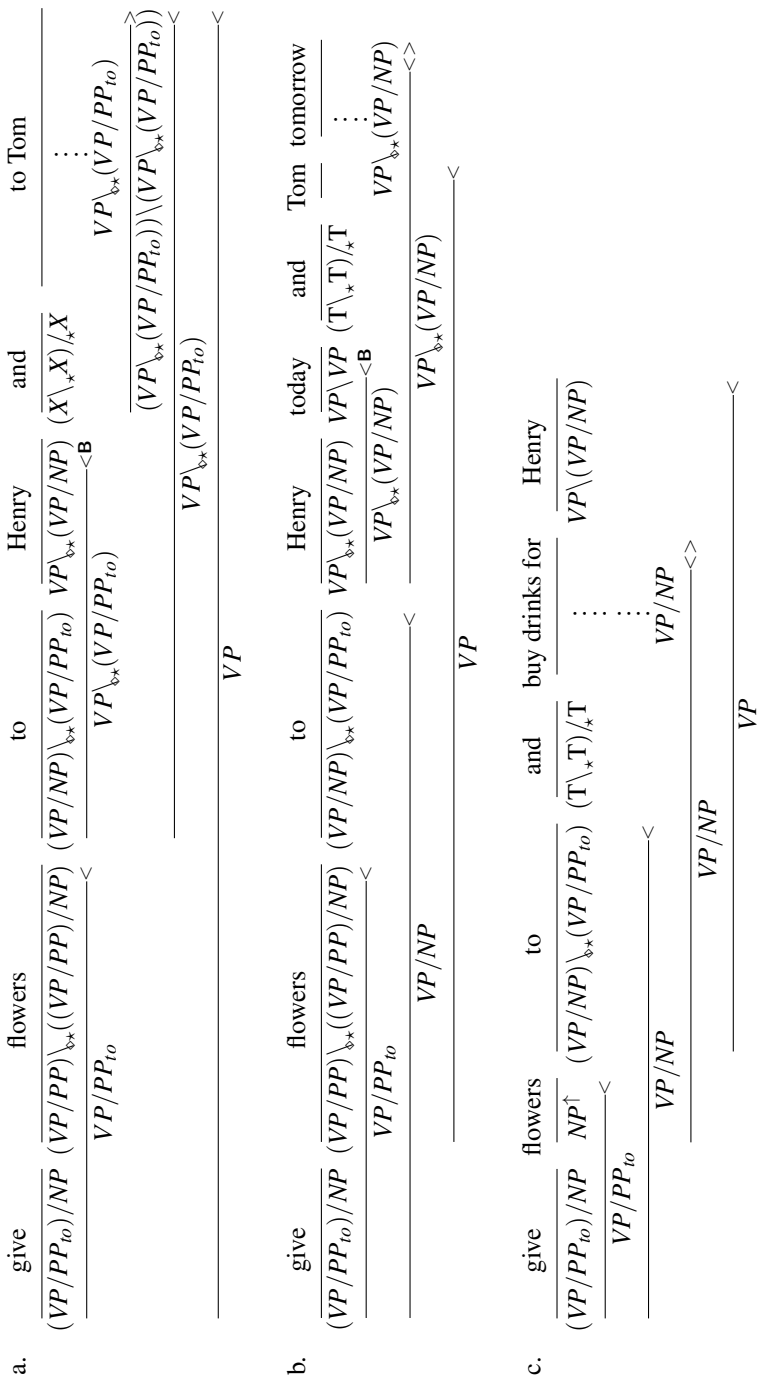
a.
give
flowers
to
Henry
and
to Tom
b.
give
flowers
to
Henry
today
and
Tom
tomorrow
c.
give
flowers
to
and
buy drinks for
Henry

Figure 10.3:

a.

| take | the BUS | TO | and | the TRAIN FROM | 30th St. station |
|---|---|---|---|---|---|
| $(VP/PP_{loc})/NP$ | $NP^{\uparrow}$ | $(VP/NP)\backslash_{\diamond\star}(VP/PP_{from})$ | $(\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T}$ | : | $VP\backslash(VP/NP)$ |
| | $(VP/NP)\backslash_{\diamond\star}((VP/PP_{to})/NP)$ $<\mathbf{B}$ | | | $(VP/NP)\backslash_{\diamond\star}((VP/PP_{from})/NP)$ | ⋮ |
| | $(VP/NP)\backslash_{\diamond\star}((VP/PP_{loc})/NP)$ $<>$ | | | | |
| $VP/NP$ $<$ | | | | | |
| | | $VP$ $<$ | | | |

b.

| travel | TO | TOMORROW | and | FROM | the NEXT day | a very strange place |
|---|---|---|---|---|---|---|
| $VP/PP_{loc}$ | $(VP/NP)\backslash_{\diamond\star}(VP/PP_{to})$ | $VP\backslash VP$ | $(\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T}$ | $(VP/NP)\backslash_{\diamond\star}(VP/PP_{from})$ | $VP\backslash VP$ | $VP\backslash(VP/NP)$ |
| | $VP\backslash_{\diamond\star}(VP/PP_{to})$ $\mathbf{B}_{\times}{}^{2}$ | | | | ⋮ | |
| | $VP\backslash_{\diamond\star}(VP/PP_{loc})$ $<>$ | | | | $VP\backslash_{\diamond\star}(VP/PP_{from})$ | |
| $VP/NP$ $<$ | | | | | | |
| | $VP$ $<$ | | | | | |

Figure 10.4:

particle-like category for prepositions proposed for English in (65) of chapter 9.4, but that it is restricted to full NP arguments using the same feature that was used in that chapter to exclude them for English extractable complement subjects, as in the following:

(31) le := $(VP\rfloor_{\diamond\star}NP_{-WH})\rangle_{\diamond\star}(VP/PP_{le})$

McCloskey further (n.4) claims that right-node-raising preposition stranding in Irish and a number of languages showing the same asymmetry is unlike English in only being possible if *both* conjuncts include a stranded proposition. This too can be captured if $VP\rfloor_{\diamond\star}NP_{-WH}$ is marked for non-*wh*-government can only coordinate with a similarly restricted category.

McCloskey's observation should therefore not be taken as evidence against the present claim that formation of the residues of leftward and rightward movement involve the same derivations and types.

## 10.6 Multiple and Correlative Coordination

The formal basis for multiple coordinations like the following have remained an irritatingly unsolved problem for all theories of coordination (Zwart, 2005; Lasnik, 2011; Chomsky, 2013):

(32) a. I met [Bob, (and) Carol, (and) Ted . . . , *(and) Alice].
b. I [marinated, (and) cooked, (and) . . . , *(and) ate] the mushrooms that I gathered beneath the gloomy ramparts of the Grange.
c. He is either a fool, (or) a madman, . . . , *(or)/*and a crook

The category in (2), $(\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T}$, only accounts for case where the optional conjunctions are all present. However, in "monosyndetic" (infixing) coordination in English and many other languages only the final conjunction is obligatory. (In many languages, including Japanese and Thai, the final conjunction can also be omitted, but only when it is equivalent to "and".)[9]

One might be tempted to assume that all the intermediate conjuncts such as *Hardy* in (32a) bear the same category as as the final *and Willis*, either by virtue of the comma or the related phonological marking bearing the conjunction category, or (more likely, since the commas and intonational markers are frequently dropped) by virtue of a unary rule. (This is the approach taken by Maxwell and Manning 1996 within an LFG framework.) However, the same mechanism would allow the final conjunction to also be dispensed with, contrary to observation.

The disjunctive agreement between *either* and *or/*and* in (32c) shows that such "correlative" coordination or "conjunction doubling" (Dik, 1968; Progovac, 1998) has to

9. In "bisyndetic" (postfixing) coordination of the kind found in some Athabaskan languages each conjunct seems obligated to carry a conjunction (Haspelmath, 2004:5)

mark the type of conjunction (Zhang, 2007). We will therefore replace the earlier conjunction types with categories like the following:[10]

(33) and $:= (\mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{+})/_{\star}\mathrm{T} : \sqcap$
or $:= (\mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{+})/_{\star}\mathrm{T} : \sqcup$
...
both $:= \mathrm{T}/_{\star}\mathrm{T}_{and} : \lambda x.x$
either $:= \mathrm{T}/_{\star}\mathrm{T}_{or} : \lambda x.x$
...

These categories combine with left conjuncts by the following special-purpose versions of the backward application rule:[11]

(34) *Backward Application: Conjunction Rules*

$$
\begin{array}{llll}
\mathrm{T} : a & \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{+} : f & \Rightarrow_{<+} & \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{\star} : \lambda x.f\,a \sqcap x \\
\mathrm{T} : a & \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{+} : f & \Rightarrow_{<+} & \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{\star} : \lambda x.f\,a \sqcup x \\
\ldots & & & \\
\mathrm{T} : a & \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{*} : f & \Rightarrow_{<\star} & \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{\star} : \lambda x.f\,a \sqcap x \\
\mathrm{T} : a & \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{*} : f & \Rightarrow_{<\star} & \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{\star} : \lambda x.f\,a \sqcup x \\
\ldots & & & \\
\mathrm{T} : a & \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{*} : f & \Rightarrow_{<} & \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{\star} : f\,a \\
\mathrm{T} : a & \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{*} : f & \Rightarrow_{<} & \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{\star} : f\,a \\
\ldots & & & \\
\mathrm{T} : a & \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{*} : f & \Rightarrow_{<} & \mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{+} : f\,a \\
\mathrm{T} : a & \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{*} : f & \Rightarrow_{<} & \mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{+} : f\,a \\
\ldots & & &
\end{array}
$$

The first two sets of rules ensure that there is at least one left conjunct, while the second two sets of rules terminate the iteration.

Thus, we allow derivations like the following:

10. *neither... nor* can be treated similarly, but are omitted here because they involve the complication of negative polarity. The categories given here do not restrict *both... and* to binary conjunction, allowing *?both Bob, Carol, Ted, and Alice*. Further restricting them in this way is suggested as an exercise.

11. Languages that omit the final conjunction differ only in having an additional unary rule $\mathrm{T} : a \Rightarrow NP^{\uparrow}_{and}\backslash_{\star}NP^{\uparrow +} : a\sqcap$.

(35)
$$
\begin{array}{lllllll}
\text{either} & \text{Bob,} & \text{Carol,} & \text{Ted,} & \text{or} & & \text{Alice} \\
\overline{\mathrm{T}/_{\star}\mathrm{T}_{or}} & \overline{NP^{\uparrow}} & \overline{NP^{\uparrow}} & \overline{NP^{\uparrow}} & \overline{(\mathrm{T}_{or}\backslash_{\star}\mathrm{T}^{+})/_{\star}\mathrm{T}} & & \overline{NP^{\uparrow}} \\
 & & & & \multicolumn{2}{c}{\overline{NP^{\uparrow}_{or}\backslash_{\star}NP^{\uparrow +}}\;>} & \\
 & & & \multicolumn{3}{c}{\overline{(NP^{\uparrow}_{or}\backslash_{\star}NP^{\uparrow \star})}\;<+} & \\
 & & \multicolumn{4}{c}{\overline{(NP^{\uparrow}_{or}\backslash_{\star}NP^{\uparrow \star})}\;<*} & \\
 & \multicolumn{5}{c}{\overline{NP^{\uparrow}_{or}}\;<} & \\
\multicolumn{6}{c}{\overline{NP^{\uparrow}_{or}}\;>} &
\end{array}
$$

The penultimate conjunction step of this derivation can iterate arbitrarily, and multiple conjunctions are allowed, so long as the final conjunction is included, to yield the following alternatives and contrasts:

(36) a. Bob, Carol, Ted, and Alice
b. Bob and Carol, Ted and Alice
c. Bob and Carol, and Ted and Alice
d. *Bob, Carol, Ted, Alice
e. *Bob, and Carol, Ted, Alice

As Winter (2001, 2007) points out, when intermediate conjunctions are present, there are multiple possibilities for distributivity across conjuncts, but when only the final conjunct is included, distributivity is fixed. This is predicted under the present account.

The fact that conjunctions and disjunction of arguments are type raised and marked for their conjunction type means that we immediately account for the fact that the categories given earlier for the *both* and *either* can "float" to higher positions in the sentence than the conjunction itself (Larson, 1985; Zhang, 2008):

(37) a. Kim married either Bob, Carol, Ted, or Alice.
b. Kim either married Bob, Carol, Ted, or Alice.

On the assumption that type-raised categories such as $itNP^{\uparrow}_{and}$ mark their result for the same conjunction type, we have derivations like the following for (37b)

(38)
$$
\begin{array}{lllll}
\text{Kim} & \text{either} & \text{married} & \text{Bob, Carol, Ted, or Alice} & \\
\overline{NP^{\uparrow}} & \overline{\mathrm{T}/_{\star}\mathrm{T}_{or}} & \overline{(S\backslash NP)/NP} & \overline{(S\backslash NP)_{or}\backslash((S\backslash NP)/NP)} & \\
 & \multicolumn{3}{c}{\overline{(S\backslash NP)_{or}}\;<} & \\
 & \multicolumn{2}{c}{\overline{S\backslash NP}\;>} & & \\
\multicolumn{3}{c}{\overline{S}\;>} & &
\end{array}
$$

Since this is exactly the same mechanism used in other chapters for *in situ wh*-movement, clitic-climbing, and scope of quantifiers, negation, focus particles, and intonational phrasing, it is immediately predicted to be, like those constructions, both unbounded and sensitive to island constraints, as noted by Larson:

(39) a. Kim either [said that he would marry]$_{(S\backslash NP)/NP}$ [Bob, Carol, Ted, or Alice]$_{NP^{\uparrow}_{or}}$.

b. #Kim either [met a man who married]$_{\#}$ Bob, Carol, Ted, or Alice.

As in all those constructions, the similar intonation and information-structural requirements of correlative coordination are determined in the same way. However, correlatives like *either* are neither quantifiers (Dougherty, 1970; Stockwell, Schacter, and Partee, 1973), initial conjunctions, (Ross, 1967; Gazdar, 1981; Gazdar et al., 1985), nor focus particles (Hendriks, 2004; Johannessen, 2005; Zhang, 2008) (see Zhang for a historical and cross-linguistic review of these accounts).

From now on we will revert to the simpler category $(\mathrm{T}\backslash_{\star}\mathrm{T})/_{\star}\mathrm{T} : \sqcap$ as an abbreviation for full categories like $(\mathrm{T}_{and}\backslash_{\star}\mathrm{T}^{+})/_{\star}\mathrm{T} : \sqcap$, and will regard the simple backward application rule as abreviating the conjunction rules (34), except in cases of actual multiple conjunction.

## 10.7 Gapping and Stripping

Steedman 1990 and *SP* show that it is also predicted that head-medial languages and constructions like the following example of the English "gapping" construction universally pattern with the head-initial constructions in having the non-standard constituent to the right:

(40) I like Ike and Chester, Adlai.

Unlike cluster coordination, a further discourse-based anaphoric process is required to recover the gap itself in example like (40). Such anaphoric processes are notoriously tolerant of minor feature mismatches such as agreement in the gapped material. Compare

(41) *I like Ike and Chester like Adlai.

The mechanism for gapping outlined in *SP* and extended in Steedman (2017) offers a second "gapping" derivational route for cluster coordinations like (23), according to which residues like *I showed* are recovered by this further process. Since as we have seen, gapping is notoriously less sensitive to syntactic detail than constituent coordination, this may explain the fact that examples like the following, in which the types of the cluster conjuncts require different diathesis alternates of the verb *show*, do not seem too bad, as Beavers and Sag (2004) point out:

(42) I showed [three boys a movie]$_{(S\backslash NP)\backslash(((S\backslash NP)/NP)/NP)}$, and [a video to two girls]$_{(S\backslash NP)\backslash(((S\backslash NP)/PP)/NP)}$.

If so, such examples do not controvert the generalization that coordination is essentially an operation over like types, contrary to their claim.

A further focus of doubts about the otherwise overwhelming generalization that coordination operates over like types arises from examples like the following:

(43) The temperature is [ninety]$_{NP}$ and [rising]$_{AP}$.

Cann et al. (2005, 216–219) discuss some more complex cases of apparent conjunction of unlike types, where there is a mismatch on a minor feature such as polarity or participial features of the VP under right-node raising.

*TS*:99–101 relates such examples to the phenomenon of right- and left-conjunct agreement in languages like Hebrew (Doron, 2000) and Welsh (Borsley et al., 2007), and suggests that conjunct dominance in agreement might better be analyzed as a low-level morphophonemic effect, like French *liaison*, the English *a/an* alternation in the indefinite article, or initial consonantal mutation in Celtic languages, including Welsh (cf. Benmamoun 2000 on agreement in Arabic as PF process).

For present purposes, we can ignore such minor details of exactly what counts as type compatibility for the purposes of coordination. The important generalization is that coordination is an operation over like types, however that is defined.

## 10.8 On the distinction between Across-the-Board and Parasitic Extraction

There is a natural tendency to see an analogy between across-the-board extraction under coordination and the "parasitic" extractions from adjuncts considered in the previous chapter, as Ross (1967) noticed, and as Williams (1990), Levine and Hukari (2006), Zhang (2010), and Chaves (2012) have proposed. Both involve a single extracted element with multiple dependencies, and therefore present a problem for the simplest movement accounts. It is inevitable that there should be some further parallels, because both of them involve an extraction from a right-adjunct from which extraction is normally disallowed.

(44) a. articles that$_i$ I [filed$_i$ [without [reading$_i$]]]
b. articles that$_i$ I [filed$_i$ [and [forgot$_i$]]]

However, in other respects, the two constructions look very different. (Engdahl, 1983; Munn, 1993; Postal, 1993). Most important, as was seen in (25b) of chapter 9.4 the "true" extraction in (44a) is allowed without the "parasitic" gap in the adjunct, whereas *neither* extraction in ATB can occur without the other.[12]

Second, as Postal points out (among numerous other arguments), the types that can parasitically extract are much more restricted that those that can do so under the ATB condition, essentially to just NP and subcategorized PP.

Thus we have

(45) a. *How sick$_i$ did John look$_i$ without actually feeling$_i$?
b. How sick$_i$ did John look$_i$ and actually feel$_i$?

12. See section 10.1 on claimed exceptions

(46) a. *How$_i$ did John mend$_i$ the car before fixing$_i$ the truck?
b. How$_i$ did John mend$_i$ the car and fix$_i$ the truck?

The reason is that, as is well known, the way adjuncts extract is quite unlike that for arguments. In fact, such extractions are not in general unbounded, as we saw in section 9.7:

(47) a. How$_i$ did John mend$_i$ the car?
b. *How$_i$ do you doubt that John mended$_i$ the car?

While Hornstein and Nunes (2002) offer an explanation in terms of the violation of the principle of Last Resort for why extraction of adjuncts is limited in this way, this principle then has to be "amnestied" by a Parallelism Constraint on coordination acting as a bare output constraint. That seems to mean that ATB remains an explicit constraint, contrary to Minimalist principles.

In this connection, it is worth recalling that the **B**- and **S**-combinatory rules of chapter 9 that permit extraction only apply to arguments Z. Coordination is mediated by a quite different mechanism limited to like-typable constitents, It therefore allows the following facts to be captured without further stipulation:

(48) a. How$_i$ did John mend$_i$ the car and fix$_i$ the truck?
b. *How$_i$ do you doubt that John mended$_i$ the car and regret that he fixed$_i$ the truck?

Because we have lexicalized the bounded constructions like raising and control, they interact correctly with coordination. Thus in the following, the controled subject of the right-node-raised infinitival is separately bound by the logical forms of the two conjuncts:

(49) a. Lester agreed and Chester refused to leave.
b. Chester seemed and Lester appeared to be drunk.

## 10.9 Discussion

Many of the supposed asymmetries between leftward and rightward outlined above depend upon the analysis of coordination. The predominant approaches within the minimalist program seem to be based either on the idea of sideward movement (Nunes, 2004) or the related idea of multidominance (McCawley, 1982; Goodall, 1983; Lin, 2002; Sarkar and Joshi, 1996; Wilder, 1999, 2008; Chen-Main, 2006; De Vries, 2009; Citko, 2011). More recently, there has been a return to the idea of movement and/or deletion under identity (Hirsch, 2017; Torr and Stabler, 2016; Schein, 2017), partly because of difficulties for multidominance arising from the interaction of quantifier scope and coordination, to which we turn in a later chapter. Zhang (2010) proposes yet another approach, based on pronominal anaphora.

However, all of these approaches require structural parallelism conditions on coordi-

nate structures, and none of them seems to explain Ross's 1970 observations concerning the relation of the direction of gapping to canonical word-order, one of the strongest syntactic universal generalizations that is known.

In contrast, CCG defines parallelism in terms of simple identity of syntactic types, and a syntax that eliminates all operations equivalent to overt or covert "movement", "deletion", "copying", and all other varieties of action-at-a-distance, in favor of type-dependent combinatory syntactic and semantic merger over elements that are overt, adjacent, and directly interpretable, in which Ross's generalization is explained as a corollary of the Combinatory Projection Principle, which says that information specified in the lexicon, including directionality, must be projected by syntactic derivation, an assumption akin to the inclusiveness condition.

As a consequence, CCG differs from the GPSG account of coordination (Gazdar, 1980, 1981; Gazdar et al., 1985). Its use of directional types makes right node raising and right cluster coordination from head-initial verbs a prediction of the theory. GPSG requires an explict rightward displacement rule $\alpha \quad \rightarrow \quad \alpha/\beta \quad \beta$ to license rightward extraction (Gazdar, 1981:178-9).[13]

The accounts that CCG affords for *wh*-extraction and right node-raising dependencies depend on an identical mechanism of composition and case (type-raising), making the intervening material into a typable constituent. There is therefore a strong prediction of symmetry between the two constructions, such that whatever can undergo extraction can also undergo right node-raising, and vice versa.

Apparent exceptions to the symmetry of left- and right-extractions seem to reflect subtle interactions of the direction of extraposition with information-structural aspects of semantics, such as theme and rheme status, that are entirely independent of the syntactic mechanism of extraction, rather than revealing any shortcoming in the central assumption of CCG that rightward and leftward extraction are syntactically symmetric as a consequence of the Combinatory Projection Principle and the Projective Dependency Principle of chapter 2.

**Exercise** : Try to come up with an alternative explanation for rightward-only prepositions stranding in Irish in McCloskey's example (30)

**Exercise** : The above account of coordination requires us to radically rethink our definition of the notion "constituent". Defend (or attack) the CCG notion of constuency as combinatory typability against (or with) the traditional linguistic notion(s).

13. This rule is dropped in Gazdar et al., 1985, which does not include any analysis of right node raising or Heavy NP Shift.

# Chapter 11

# Combinatory Minimalism

There will be seen in it demonstrations of those kinds which do not produce as great a certitude as those of Geometry, and which even differ much therefrom, since whereas the Geometers prove their Propositions by fixed and incontestable Principles, here the Principles are verified by the conclusions to be drawn from them; the nature of these things not allowing of this being done otherwise
—*Treatise on Light* Christiaan Huygens, 1690 (tr. Silvanus P. Thompson)

The present account eliminates from the formal theory of grammar every form of movement, whether of the overt-, covert-, A-, Ā-, *wh*-. quantifier-, scrambling-, sideward-, remnant-, head-, topic-, or focus- variety. As a consequence, it necessarily also renders unnecessary many constraints and minimality conditions limiting movement.

This reduction is achieved by radical lexicalization—that is by specifying all the long-range dependencies that have been described in terms of movement via lexical entries for their functional governors (such as control verbs, quantifier determiners, and relative pronouns) with a syntactic type (such as $(VP/VP_{to})/NP$, $(S/(S\backslash NP))/N$, or $(N\backslash N)/(S/NP)$), and a logical form (such as $\lambda x \lambda p.persuade\,(p\,x)\,x$,, $\lambda n \lambda p.p\,(skolem\,n)$, or $\lambda p \lambda n \lambda x.n\,x \wedge p\,x$) that already expresses the (possibly unbounded) dependency statically, via $\lambda$-bound variables.[1]

These lexical entries express the entire combinatoric potential of the word-category pair, including subcategorization if it is a verb, and such semantic relations between its arguments as control—again, statically. In the terms of the competence theory, a subset of these lexical types corresponding to arguments are also *cased* or type-raised, either statically, via inflection as in Latin, or dynamically or "structurally" as in English, or by a "quirky" mixture of the two, as in Icelandic.

This means of course that many conditions on the grammar are recast as conditions on the lexicon, where they must be analysed as part of the definition of "possible lexical category". For example, subjacency must be explained in terms of the CCG type-system that is investigated in section 11.5 below, which universally excludes non-subjacent "super-raising" and "super-control" verbs from the theory of grammar.

1. A combinatory variable-free notation for logical form is possible, but is, as noted earlier, considerably less readable.

However, the generalization in question can now be interpreted as applying to the type-system of logical form, and ultimately to the underlying conceptual representational "language of mind" that makes language and language acquisition possible, as proposed in chapter 2 and in appendices B and A.

The lexical entries of the radicalized lexicon are projected syntactically by a small set of combinatory rules of an even smaller fixed set of types confined to (besides application), composition and substitution, onto constituents including sentences, bearing derived categories, again consisting of a type and a logical form, the latter being assembled entirely compositionally by the derivation.

The inclusion of morpholexical type-raising, and of syntactic composition and substitution means that the inventory of constituent types onto which the lexicon is projected by combinatory derivation is extended greatly beyond any traditional account of constituency, so that, for example, "John says Mary found", "burn without reading" and "a policeman a flower" are all derivationally typable and interpretable constituents in their own right, rather than being of type S or VP containing the residues of movement or deletion.[2]

The following generalization then follows:

(1) *Coordination, relativization, quantifier-scope-taking, and intonation structure are all defined over the same generalized notion of syntactic constituency, which is entirely transparent to a surface compositional semantics. All of these constructions obey the constituent condition on rules directly, without requiring movement, deletion, or mediating structural representations.*

However, we also noticed in chapter 4.1 that, in order to achieve this generalization, we had to give up some traditional assumptions concerning constituency in surface syntactic derivation, and admit fragments like "I met", "I think I met", and "Mary a book last Saturday" as syntactically typable constituents and first-class citizens of the grammar, and to live with the idea that even a simple transitive sentence has a surface derivation in which the object c-commands the rest of the derivation. This raises the question of how we can define processes which have been argued to depend on traditional constituency and structural relations of c-command, of which the two most important are anaphora and coreference on the one hand, and variable binding by operators such as Universal Quantifiers on the other. These topics are the subject of further chapters in Part III below on semantics and anaphora. However, the following remarks are in order at this point.

2. This departure from more traditional notions of constituency is sometimes held against CCG as evidence of inconsistency as a theory of grammar (e.g. Osborne and Groß, 2016). It is actually evidence of the inconsistency of all of the traditional tests for constituency *except for coordination and intonation*.

## 11.1 Binding and Coreference

First, neither phenomena can depend on surface constituency, since there are right node-raising derivations which force a possibly embedded argument to take derivational scope over the residue, without forcing either a violation of the anti-c-command condition on pronouns and antecedents or a wide-scope-only reading in the object, contrary to the facts:

(2) a. $\text{Lola}_{S_t/(S/NP)}$ [Anna thinks Manny likes]$_{S/NP}$.
b. [$\text{Anna}_i$ thinks Manny likes]$_{S/NP}$ $\text{HER}_i$ $_{S\backslash(S/NP)}$.

(3) a. Some $\text{saxophonist}_{S_t/(S/NP)}$, [every woman likes]$_{S/NP}$. $\forall\exists/\exists\forall$
b. [Every woman likes]$_{S/NP}$ some $\text{saxophonist}_{S\backslash(S/NP)}$. $\forall\exists/\exists\forall$

(It was partly for this reason that we adopted a lexicalized clitic analysis of reflexives and reciprocals in chapter 7.)

However, traditional notions of constituency and c-command still hold in CCG at the level of logical form, which is arguably the level of representation that those traditional intuitions concern in the first place. To the extent that they are grammatical phenomena at all, both anaphoric coreference and scope-taking must therefore be defined at the level of logical form, as they are under the program of Chomsky (1995b).

Nevertheless, it is also well-known from examples like the following, repeated from the introduction, that binding is not solely determined in any simple sense by structural command at any level (cf. Reinhart, 1983a; Higginbotham, 1983; Lebeaux, 1991, 2009; Jacobson, 1994; Hornstein, 1995:108-111; Romero, 1998, Fox, 1999, Büring, 2005:256-260, Takahashi and Hulsey, 2009, Barker, 2012, and Bruening, 2014):

(4) The person that $\text{Lola}_i$ works for $_j$, $\text{she}_i$ $\text{likes}_j$.
Which person that $\text{Lola}_i$ works for $_j$ does $\text{she}_i$ $\text{like}_j$?

The observation has led to proposals for "Late Merge", free type-lifting/continuation-passing, "Roll-up" movement, or definition in terms of "Almost C-command", higher-type $((e \rightarrow t) \rightarrow t)$ traces, and/or "hybrid" movement systems involving both copies and traces (Lechner, 1998, 2018; Keine and Poole, 2018).

Instead we will follow Bruening (2014) in assuming that coreference is not a relation between referring expressions represented by nodes in a logical form, but rather a relation between referring expressions and referents in a dynamically-changing contextual model. The reason that "Lola" and "she" can co-refer in (6) is that by the time "she" is processed, the referent of "Lola", namely an entity *lola*, has been added to the contextual model as a possible antecedent.

The examples themselves in (6) show that the process of updating the contextual model cannot wait until the end of the sentence. However, it cannot just be as soon

as the referring expression "Lola" is complete, as that would allow examples like the following, contrary to Principle B:

(5) *Lola likes her.

It must rather be that the entity *lola* and any other referents that have been mentioned in the same binding domain gets added to the contextual model as soon as either (i) the domain itself is exited from, or; (ii) a subordinate new binding domain is entered,

So, for example, in addition to examples (6), the following are possible where $^{\uparrow lola}[$ and $]^{\uparrow lola}$ represent type i and type ii updates, and $^{\downarrow her}$ represents reference:

(6) a. Lola$_i$ thinks $^{\uparrow lola}$[she$_i^{\downarrow her}$ is a genius].
b. [That Lola$_i$ is a genius]$^{\uparrow lola}$ surprises her$_i^{\downarrow her}$
c. [Pictures of Lola$_i$]$^{\uparrow lola}$ amuse her$_i^{\downarrow her}$.

There is a little more to say about the examples in (31) of chapter 1, repeated here:

(7) a. [The person that she$^{\uparrow she}$ works for]$_j$, Lola likes$_j^{\uparrow lola}$.
b. [Which person that she$^{\uparrow she}$ works for]$_j$ does Lola like$_j^{\uparrow lola}$?

While sometimes misleadingly referred to as "backward anaphora", what is really going on in these examples is that when a pronoun without an available antecedent in the contextual model (here, "she"), is encountered, a proxy referent (here, *her*), with the sole property of being "given", or in the terms of chapter 6 "background", and the logical form of a subsequent definite (i.e. given) NP (here, "Lola") is then predicated of it, making it refer definitely (here, to *lola*). Note that "backward anaphora" to indefinites is degraded:

(8) a. #Which person that she$_i$ works for does some woman$_i$ like?
b. Which person that a woman$_i$ works for does she$_i$ like?

The mechanism for modifying the referential context in this way is located in the lecxical logical forms of the heads of clauses and NPs. For example:

(9) Lola $:= NP^{\uparrow} : \lambda p.p\,lola$
thinks $:= (S\backslash NP)/S : \lambda s\lambda y.thinks\,s\,y \wedge push\,y\,context$
she $:= NP^{\uparrow} : \lambda p.her = pop\,context \wedge p\,her$

(10)
$$\dfrac{\dfrac{\dfrac{\text{Lola}}{\begin{array}{c}NP^{\uparrow}\\ :\lambda p.p\,lola\end{array}} \quad \dfrac{\text{thinks}}{\begin{array}{c}(S\backslash NP)/S\\ :\lambda s\lambda y.thinks\,s\,y \wedge push\,y\,context\end{array}} \quad \dfrac{\text{she}}{\begin{array}{c}NP^{\uparrow}\\ :her = pop\,context \wedge \lambda p.p\,her\end{array}}}{\begin{array}{c}S/S : \lambda s.thinks\,s\,lola \wedge push\,lola\,context\\ \mathbf{B}\end{array}} \quad \dfrac{\text{is a genius}}{\begin{array}{c}S\backslash NP\\ :\lambda y.genius\,y\end{array}}}{\begin{array}{c}S : thinks\,(genius\,lola)\,lola\end{array}}$$

(Note that unlike most derivations in the book, this derivation depends on left-right incrementality.)

The notion of binding domain invoked above resembles the minimalist notion of "phase" (Chomsky, 2001). However, it is an exclusively semantic notion, rather than a syntacxtic one. It follows that no "Phase Impenetrability Condition" (PIC) is needed or can be defined.

## 11.2 Grammar without Action-at-a-Distance

The driving force behind CCG was from the very first to reduce transformational action-at-a-distance to operations over string-adjacent contiguous elements. This was enshrined in the earliest papers in the form of a Principle that can be stated as follows (cf. Ades and Steedman, 1982:533, *SP*:54).

(11) *The Principle of Adjacency*
Syntactic rules can only apply to pairs of string-adjacent typed categories.

Since combinators are by definition operators that apply to adjacent terms, and since we know from the work of Schönfinkel (1924), Curry and Feys (1958), and Smullyan (1985, 1994) that even quite small collections of unrestricted combinators have the full expressive power of the $\lambda$-calculus, the principle of Adjacency says little more than that a combinatory calculus is adequate to express natural language grammar including the language of logical form, and that action-at-a-distance is not a formal necessity, *contra* Chomsky (2005):10.

However, as we have seen, combinatory categorial grammar is much more restricted than the cominatory calculi that are equivalent to the $\lambda$-calculi. First, the categories are *typed*, and those types define linear precedence. Second, the syntactic combinatory rules that project precedence information are constrained by the Combinatory Projection Principle (5) of chapter 2, which is itself a corollary of the Principle of Adjacency (11) and the following two more specific principles (cf. *SP*:54) in which we continue to distinguish that argument functor of a combinatory rule whose range X is the range of the result of the rule as the "governing functor", and that argument functor whose range Y is that of one of the arguments of the principal functor as the dependent functor:

(12) *The Principle of Linear Consistency*
If the linear precedence specified for the argument of the governing functor corresponding to the result of the dependent functor is rightward (leftward), then the dependent functor must be adjacent to the right (left).

(13) *The Principle of Linear Inheritance*
The linear precedence specified for any argument of the input functors that appears in the result of a combinatory rule must be the same as that specified on the input functor type(s).

These three principles collectively constitute what in earlier chapters was referred to as the Combinatory Projection Principle (CPP). They constrain the expressive power of CCG, putting it in a class of grammars that is the least more expressive than context-free that is of linguistic interest. We shall see that, in the terms of the Chomsky hierarchy of natural families of language, it is far less expressive that the context-sensitive (Type 1) languages, and far smaller than Joshi's 1985 subclass of "mildly context-sensitive" languages, so it seems appropriate to call it "nearly context-free".

## 11.3 The Place of CCG in an Extended Language Hierarchy

Vijay-Shanker, Weir, and Joshi (1987), Weir (1988), Joshi et al. (1991), and Vijay-Shanker and Weir (1994) showed in a series of papers that, under the formalisms that were then current, four grammatical formalisms—TAG, CCG, Linear Indexed Grammars (LIG, Gazdar, 1988) and Head Grammars (HG, Pollard, 1984) were weakly equivalent—that is, that they could capture the same set of languages or stringsets. They also showed that the worst-case complexity of the decision problem for such grammars—that is, the problem of deciding whether a string was in the language or not—was polynomial, a property which has important consequences for efficient parsability. They did this by showing that such grammars were drawn from a "natural family of languages" (AFL), characterized by a class of phrase-structure rules, an automaton, and exemplars of languages that are not recognizable by any less powerful grammar or automaton.

That is not to say that these grammars are strongly equivalent. In particular, their derivation trees and their treatments of unbounded dependencies are quite different (although *SP* presents an intuitive demonstration that for every CCG there is a strongly equivalent LIG).

Vijay-Shanker and Weir went on to show that the natural generalization of TAG was to the class of linear context-free rewriting systems (LCFRS), which under some further restrictions are polynomially decidable (Satta, 1992). LCFRS are equivalent as a class to multiple context-free grammars (MCFG, Seki, Matsumura, Fujii, and Kasami, 1991).

### 11.3.1 The Extended Language Hierarchy

These observations require us to extend the Chomsky Language Hierarchy by interpolating a number of levels besides the original four types, as in figure 11.3.1, in which lowercase letters represent terminals, uppercase leters represent nonterminals, lowercase greek letters represent strings of terminals and nonterminals, and exponents on

terminal symbols as in $a^n$ represent sequences of $n$ occurrences of that terminal.

| Grammar Type | Automaton | Rule-types | Exemplar |
|---|---|---|---|
| **Type 0**: RE | Universal Turing Machine | $\alpha \rightarrow \beta$ | PA-valid |
| **Type 1**: CSG | Linear Bounded Automaton (LBA) | $\phi A \psi \rightarrow \phi \alpha \psi$ | $a^{n!}$ |
| MCFG (LCFRS) | $i$th-order EPDA | $A_{[[(i),\ldots]\ldots]} \rightarrow \phi B_{[[(i),\ldots]\ldots]} \psi$ | $\mathscr{P}(a^n b^n c^n)$ |
| IG | Nested Stack Automaton (NSA) | $A_{[(i),\ldots]} \rightarrow \phi B_{[(i),\ldots]} \psi C_{[(i),\ldots]} \xi$ | $a^{2^n}$ |
| LIG/CCG/TAG | Embedded PDA (EPDA) | $A_{[(i),\ldots]} \rightarrow \phi B_{[(i),\ldots]} \psi$ | $a^n b^n c^n$ |
| **Type 2**: CFG | Push-Down Automaton (PDA) | $A \rightarrow \alpha$ | $a^n b^n$ |
| **Type 3**: FSG | Finite-state Automaton (FSA) | $A \rightarrow \begin{cases} \text{a B} \\ \text{a} \end{cases}$ | $a^n$ |

Figure 11.1: The extended language hierarchy

As in the original hierarchy introduced in chapter ??, each level properly contains the level below, except that IG and LCFRS are mutually properly intersecting.

The fact that the original Chomsky hierarchy had four levels should not lead one to assume that those levels are in any computational sense equidistant. The type 0 languages are those that are recursively enumerable sets, which for present purposes means all sets with a formal definition. Such sets are explanatorily vacuous as a theory of natural language in the sense that they place no restrictions on possible languages in the sense of stringsets. Savitch (1987) shows that in this sense, type 1 (CS) grammars are not significantly less expressive than Type 0.[3]

The unnumbered overlapping LCFRS and IG classes are much less expressive than Type 1, the context-sensitive class. Nevertheless, they include many languages that seem to have highly unnatural properties, like Bach's MIX language $\mathscr{P}(a^n b^n c^n)$, consisting of all permutations on the same number $n$ of occurrences of some number of terminals, and the "non-constant growth" language $a^{2^n}$.[4]

The unnumbered level comprising CCG, TAG, LIG and HG is much less expressive

3. The restriction implicit in the linear bounded automaton of the type 1 level is merely to a Turing machine with bounded memory—that is, to something that is more like a real computer.

4. There are recurring claims in the literature that the phenomenon of reduplication, which is quite widespread in natural languages, is productive and generates subsets of strings of the form $a^{2^n}$. An early example from Manaster-Ramer (1986) concerns reduplication in certain US English dialects, which is a marker of dismissive emphasis, as in the old joke about a mother who is told that her son has an Œdipus complex, and replies "Œdipus Schmœdipus, what does it matter if he's a good boy and loves his mother?". This construction might seem to open up the possibility of rejoinders such as "#Œdipus Schmœdipus SCHMŒDIPUS Schmœdipus, your son needs help." However, native speakers invariably reject such examples, saying that such markers can only apply once only to *bona fide* lexical items.

than LCFRS, and excludes such languages. We will refer to this natural family of languages as "near- context-free". It implies the extended version of the Chomsky and Schützenberger (1963) hierarchy in figure 11.3.1.

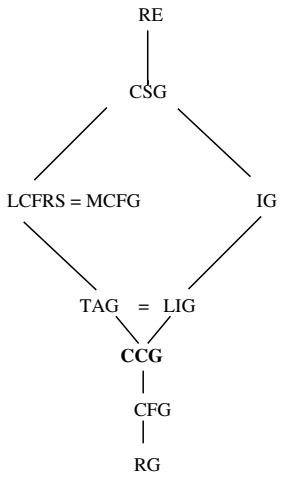


Figure 11.2: The Revised Extended Chomsky Hierarchy

This result is surprising when one reflects that the natural generalization of TAG is to the LCFRS class, while the natural generalization of LIG is to IG, so that their weak equivalence seems almost accidental.[5]

5. The definitions of CCG used in the present book is different from those in Steedman (1996) and used by Vijay-Shanker and Weir (1994) as the basis for the Joshi et al. (1991) proof of weak equivalence to TAG/LIG/HG. The main difference is that in early version of CCG including *SP*, type restrictions on the variables $X, Y, Z, \ldots$ in combinatory rules were allowed. (For example, crossed composition could be restricted in English to cases where $Y = S'$, excluding crossed composition of nominal adjuncts $N\backslash N$). In the present version of the theory, such type-based restrictions on rules are disallowed. Instead, slash-types do similar work in limiting overgeneration. (For example, we saw that the type $N\backslash_{\diamond\times}N$ of English nominal adjuncts prevents the the crossed composition rule from applying to them to generate Germanic orders like *#a nice at the office man*.) Nevertheless, for the purposes of Their proof, the systems are eqivalent.
Kuhlmann et al. (2015) showed that under the definitions used by Steedman and Baldridge (2011), CCG would be strictly less expressive than under the old definition. In particular, as noted earlier, the Germanic crossed dependency construction modeled by the language $a^n b^n$ with only crossing dependencies could not be accepted without also accepting some additional word-orders with adjacent non-crossing dependencies. As we have noted in section 5.2, such mixed word orders are in fact generally characteristic of the Germanic constructions that allow such dependencies, like the Zurich version of German considered there. The Dutch construction that *only* allows the completely crossing depenendencies could not be handled syntactically by the version in Steedman and Baldridge (2011) and related papers. However, the present system of slash modalities, first proposed in FUNLG following Baldridge (2002), restores weak equivalence to TAG and LIG.

### 11.3.2 Worst-Case Complexity of CCG Parsing

Vijay-Shanker and Weir (1993) also show that the TAG/CCG/LIG/HG near- context-free grammars are recognizable in the worst case in polynomial time $On^6$ for sentences of length $n$, and present full parsing algorithms with that bound, including one for CCG (Vijay-Shanker and Weir, 1990).

It is important to be clear that the real significance of this result is not the particular polynomial identified. Worst-case complexity tells us very little about average-case complexity, and the algorithm itself may not be practicable. The significance of polynomial complexity is, rather, to guarantee the applicability of simple and efficient generalizations of "divide-and-conquer" algorithms such as CKY (Cocke and Schwartz, 1970), of the kind used for compiling programming languages, including incremental algorithms, which in the average practical case may have acceptable costs in computational terms. We return to these questions in appendix C

## 11.4 Descriptive adequacy of CCG

The Combinatory Projection Principle (5) of chapter 2 limits the expressive power of the small set of combinatory rules to be "nearly context-free", either weakly equivalent to (Joshi et al., 1991), or, under the definitions used in the present work, properly contained in (Kuhlmann et al., 2015), the tree-adjoining languages. On either reckoning, this is the lowest linguistically interesting trans-context-free automata-theoretic natural family of languages that is known. Specifically, this class of grammar is very much less expressive than either indexed grammars (IG, Aho, 1968, 1969), or full linear context-free rewriting systems (LCFRS, Weir, 1988) or the equivalent multiple context-free grammars (MCFG, Seki et al., 1991). Stabler's minimalist grammars (MG, Stabler, 2011) are weakly equivalent to LCFRS/MCG (Michaelis, 2001).

Of course, low expressive power is of no interest unless the grammar is descriptively adequate—that is, capable of expressing the data cross-linguistically, supporting an adequate semantics. It is hard to prove descriptive adequacy in this strong sense, since there is no completely firm definition of exactly which phenomena must be covered syntactically or of their semantics. (We have argued for the exclusion of extraposition, VP-anaphora, and sluicing, but others may disagree). However, the theory outlined in this book and developed in different forms and at greater length in earlier publications, and has been applied to the syntactic and semantic analysis of coordination and unbounded dependency in a very wide range of languages (*SS&I*; Steedman 1985, 1990, 2000a; Steele 1990; Whitelock 1991; Hoffman 1995; Nishida 1996; Bozşahin 1998, 2002; Komagata 1999; Baldridge 1998, 2002, Trechsel 2000; Cha and Lee 2000; Park and Cho 2000; Çakıcı 2005, 2009; Hockenmaier 2006; Ponvert 2008; Ruangrajitpakorn, Trakultaweekoon, and Supnithi 2009; Boxwell and Brew 2010; Kubota 2010; Lee and Tonhauser 2010; Bekki 2010; Tse and Curran 2010; Steedman and Bozşahin 2016; Ambati 2016; Ambati, Deoskar, and Steedman 2016a.

Whether the near- context-free formalisms are fully descriptively adequate in the sense of capturing all linguistically attested phenomena is harder to demonstrate. In the case of CCG, we can make the strong prediction that if we have four elements of types $A|B$, $B|C$, $C|D$, and $D$ , then of the 4!=24 possible orders, two cannot be recognized by CCG. Four such elements are the determiner, numerator, adjective and noun in English NPs such as "these five beautiful girls", a concrete example that has been investigated cross linguistica;lly by Greenberg (1963), Hawkins (1983), Cinque (2005), Abels and Neeleman (2009), Dryer (2009), and others). The two excluded orders are the following, corresponding in English to "*Five girls these beautiful" and its mirror image "*Beautiful these girls five":

(14) *a.* $B|C \quad D \quad A|B \quad C|D$
*b.* $C|D \quad A|B \quad D \quad B|C$

The prediction is that these word orders will not be allowed for any language on types of these forms. If a construction were found in any language that allowed them, then CCG, at least in its present form, based on the combinators **B**, **S**, and **T** would not be descriptively adequate, and would be falsified.

These particular orders are indeed absent from the 14 word-orders that are attested for these particular parts-of-speech (Cinque, 2005), although the fact that eight further orders that CCG could allow are also unattested means that this does not tell us very much. Because of the Zipfian distribution over such word-orders, the sample of attested languages is simply not large enough for a strong test of this kind, unless we turn to free word-order languages.[6]

For similar reasons, two out of the twenty-four possible permutations of the four elements of the English VP "will$_{A|B}$ have$_{B|C}$ eaten$_{C|D}$ beans$_{D}$", namely those corresponding to corresponding to "*have beans will eaten" and "*eaten will beans have", are excluded by universal grammar. If either order were attested, say in a language with an auxiliary system and freer word order than English, such as Dutch or German, then CCG in the present form would be falsified. (As we saw in chapter 2, this is not in fact the case for those languages.)

It is very difficult to test these predictions, because the availability of anaphoric linkage via pro-drop and phenomena like parenthesis and extraposition means that other analyses than those afforded by strictly syntactic projection are often implicit.

For example, Haug (2017) analyses the following Latin example (Caesar *De Bello Gallico* V.i.i) as an instance of backward adjunct control of the subject of the participial adjunct *discedens ab hibernis in Italiam* ("departing from winter quarters to Italy") by the subject *Cæsar* of the main clause *Cæsar . . . imperat* ("Caesar ordered . . . "). If that analysis is correct, then the categories are as follows, on the pattern (14a), and cannot

6. The two excluded orders are in fact among those allowed under Hawkins' (1983) version of the Greenberg's 20th universal. CHECK

combine:[7]

(15) discedens ab hibernis Cæsar in Italiam imperat

| | | | |
|---|---|---|---|
| $S/S$ | $NP$ | $(S/S)\backslash(S/S)$ | $S\backslash NP$ |
| $B/C$ | $D$ | $A\backslash B$ | $C\backslash D$ |

However, while the implicit subject of such participial adjuncts is frequently coreferential with the subject of the main clause, it can be otherwise. In particular, it can refer logophorically to the speaker or source of indirect discourse, as discusssed in section 13.3, as in the following absolute:

(16) Departing from winter quarters to Italy, the sun was shining.

*De Bello Gallico* is very much from Cæsar's point of view, being written as a self-justifying report intended to be read aloud by others (Mueller, 2012:xxiii-v).

Such dangling participials should probably be regarded, like parentheticals, as not "in construction" with the main verb. [8]

A more challenging counterexample would be an attestation in Latin of the following NP word orders as alternatives to *Hæ quinque puelæ pulchræ ambulant* ("These five beautiful girls are walking"):[9]

(17) a. *Quinque puellæ hæ pulchræ ambulant.
b. *Pulchræ hæ puellaæ quinque ambulant.

Both seem very bad to my schoolboy Latin ear, but I offer them as hostages to fortune.[10]

More generally, the prediction is that for any set of $n$ categories $\{A|B, B|C, \ldots, M|N, N\}$, of the $n!$ permutations that are possible, the number that cannot be recognized by CCG is the $n$th in a series called the Large Schröeder Numbers, of which the first few members are $\{0, 2, 30, 326, 3, 234, 31, 762, 321, 244, \ldots\}$. The Large Schröder Numbers represent the number of permutations of $n$ that are non-separable, where separability is a property related to tree rotation. The Schröder numbers grow even faster than than the factorials, so that the proportion $n!/S(n)$of permutations that are non-seperable grows rapidly with $n$.

7. I assume that the PPs are adjuncts to *discedens*. type raising via case does not affect word order, so it is suppressed here.

8. In the full text, the main clause occurs in the middle of a blizzard of absolutes, extrapositions, and other parenthetical construction: *L. Domitio Ap. Claudio consulibus discedens ab hibernis Caesar in Italiam, ut quotannis facere consuerat, legatis imperat quos legionibus praefecerat, uti quam plurimas possent hieme naves aedificandas veteresque reficiendas curarent*: "with LD and AC consulted, departing from winter quarters for Italy, as he used to every year, Cæsar ordered the legates, who he put in charge of the legions, to build as many ships in the winter as they could, and to repair the old ones."

9. One needs to take care in considering such judgements that the words do carry the categories of determiner, numerator, adjective and noun—for example, that the adjective is not read instead as an extraposed NP modifier $NP|NP$ or a predicate $S|NP$, as opposed to $N|N$.

10. I am grateful to Rachel Hurley of Cardiff University for confirming (p.c.) that these two orders are indeed ungrammatical in Latin with the intended sense.

The observation of this property was first made in Williams, 2003:203-211 for his categorial calculus CAT. CAT has a standard directional categorial lexicon and rule of application, with a combinatory operation REASSOCIATE equivalent to composition, and an operation FLIP which reverses the directionality of a functor category which does the work of type raising, achieving the effect of higher-order fronting types.[11]

Williams (2003:229-234) applies a form of his calculus restricted to forward functors and forward composition to the analysis of Hungarian serial verb order discussed by Koopman and Szabolcsi (2000), which bears some similarity to the Dutch/German cross-serial verb complexes discussed in section **??**.

These two categorial accounts are therefore closely related. However, without the addition of morpho-lexical type-raising, or case, CAT will not express the coordinate structures discussed here in chapter 10.

If we renumber the original category set $A|B$, $B|C$, $C|D$, and $D$ as X, 1, 2, 3, then (14b) also corresponds to the *1-3-X-2 constraint on movement observed by Svenonius (2007) for adjuncts, an observation which led Svenonius to complex stipulations of strong features and null functional heads to limit movement in Germanic roll-up derivations such as pied-piping. Such stipulations are unnecessary in CCG, and thereby explained.

## 11.5 Explanatory Adequacy of CCG

Since an explanatory theory should by definition have the smallest possible number of degrees of freedom, the fact that CCG is also of low, near-context-free expressive power, at the lowest linguistically interesting trans-context-free automata-theoretic level that is known, means that it also has some claim to the stronger level of *explanatory* adequacy.

Explanatory adequacy has been widely but confusingly associated in the linguistic literature since *Aspects* with the provision of an account of child language acquisition for the class of grammars involved. However, a model of language acquisition can be supplied for any theory of grammar, provided the recipient is prepared to stipulate sufficient innate apparatus to cut down the search space to a manageable size (Fodor, 1966).

The simplest and best account of acquisition is the one that minimizes the amount of innate machinery that must be stipulated, because it imposes less of a burden of explanation on the theory of evolution. That theory is the one that has the fewest degrees of freedom in the first place. Thus, a theory of acquisition requires, rather than constitutes, an explanatory theory of grammar.

11. Williams incorrectly claims (2003:209) that type-raising evades the constraints on movement that are corrolaries if FLIP. However, he has failed to notice that type-raising is a morpholexical operation in CCG (Steedman, 2000b), rather than a syntactic operation, and therefore has has no effect on the exclusion of non-separable permutations in CCG.

To identify the degrees of freedom available in CCG, we must ask two more specific questions: "What is a possible lexical entry or category?"; and "What is a possible combinatory rule?".

## 11.6 On the Notion "Possible Lexical Category"

The set of lexical types that we have availed ourselves of in categories like (13) is also very restricted. It is a lot less than the set of all possible categories defined by the rule "if $\alpha$ and $\beta$ are types, then $\alpha/\beta$ and $\alpha\backslash\beta$ are types" (Such restrictions correspond to what *Aspects* identified as "Substantive" Universals, which derive from the semantics.)

We therefore need to also compare the degrees of freedom implicit in the lexical type system of CCG with those of the Minimalist Program.

The definition of the CCG lexical type-system has a number of "levels," like the movement-based type-systems of Williams (2003), and others. These levels respectively correspond to: the language of the core predicate argument structures; the language of logical from which lexicalizes raising and control relations; the language of the morpho-lexicon, which lexicalizes the work of the minimalists "head movement" (Roberts, 2001), "scrambling" (Ross, 1967), and "sideward movement" (Nunes, 2001); the language of case or type-raising; and the language of *wh*-movement.

### 11.6.1 The Type-System of Predicate-Argument Structures

We start with the language of the predicate-argument structures of (1b) and (13), which is defined as follows:

(18) *The type-system of predicate-argument structure:*
1. types have bounded valency $\leq$4.
2. $t$ and $e$ are elementary types.
3. $e \rightarrow t$, $e \rightarrow e$, $t \rightarrow t$ and $t \rightarrow e$ are types.
4. If $\alpha$ is an elementary type and $e \rightarrow \beta$ is a type then $\alpha \rightarrow (e \rightarrow \beta)$ is a type.

Here $e$ is the type of entities, standing in for a richer ontology of entities, distinguishing people, places, times, etc., and $t$ is the type of propositions, also possibly standing in for a more diverse set of types. This simplified type system allows $in_{e\rightarrow e}$, $that_{t\rightarrow e}$, $think_{t\rightarrow(e\rightarrow t)}$and $seemingly_{t\rightarrow t}$, for example.

The level of predicate-argument structure of the lexical type system corresponds to the level of thematic structure in Minimalism. To the extent that predicate-argument structure for any language is assumed to subsume expressions of a universal predicate-argument structure (UPAS) that supports language acquisition via semantic bootstrapping, although as with HPSG's ARG-ST and Simpler Syntax's "grammatical function tier" we do not assume any fixed repertoire of thematic roles (Dowty, 1991b).[12]

12. The present language-specific logical forms are in fact a proxy for UPAS. In particular, we do not assume that the predicate we write *promise* in the logical form of "promise" is actually atomic

### 11.6.2 The Type-System of Logical Form

The second level of the lexical type-system corresponds to logical forms including relations such as raising and control. We have seen that the lexical logical forms constituted via the lambda-binders in (1b) are more diverse. They allow *properties* of type $e \rightarrow t$ as arguments of raising and control verbs, as well as arguments of types $e$ and $t$. They also allow those arguments to combine in any order, regardless of lf-command relations, under the following principle:

(19) If a predicate $p$ as defined in (18) has an argument of type $t$, then it can be realized in the predicate-argument structure of a corresponding lexical VP stem as *either* (i) an free-variable argument, also of type $t$, *or*; (ii) as the application of an argument free variable of type $e \rightarrow t$ to another free-variable argument of type $e$. The latter $e$-type argument can either be unique to this predication ("raising"), or it can be one of the original arguments of $p$ ("control", "ECM").

For example, we can now have $seem_{(e\rightarrow t)\rightarrow(e\rightarrow t)}$, as well as $seemingly_{t\rightarrow t}$, and $quickly_{(e\rightarrow t)\rightarrow(e\rightarrow t)}$, as well as $quickly_{t\rightarrow t}$.

This level of the lexical type-system corresponds to "A-movement" in the minimalist framework. The fact that such relations are subject to minimality conditions such as the Minimal Link Condition, together with exemptions from such conditions for subject control (Rosenbaum, 1967) follow immediately from the fact that these relations are lexicalized as between co-arguments.

### 11.6.3 The Type-System of the Morpho-Lexicon

To specify the stem syntactic types of the lexicon as they are usually thought of by linguists, we also need a language-independent mapping between the stem elementary semantic types and syntactic argument types (generalizing over minor features such as $S'$, $VP_{to}$, etc.), where $\mapsto$ means "is the syntactic type corresponding to" :[13]

(20) $S \mapsto t$
$NP, PP \mapsto e$
$N, VP \mapsto (e \rightarrow t)$

(The mapping itself is non-essential. When written in full, it is one-to-one, so that we could use the original semantic types as syntactic types, at considerable cost to readability by mortals.)

Languages are then free to associate directional syntactic categories specifying the linear position of constituents corresponding to arguments of those types in all possible ways, with linking $\lambda$-binders to pass their values into universal predicate-argument

in UPAS.

13. $N$ and $VP$ don't actually correspond to the same semantic type $e \rightarrow t$. $VP$ really maps to a more complex type that we might think of as $e \rightarrow (r \rightarrow t)$, where $r$ is the Reichenbach/Davidsonian "event time" or "reference time". However, we use a simplified type system, for ease of reading.

structure (UPAS).

At this stage, the language-specific lexicon can further specify restrictions on lexical categories via slash-types, which limit the rules that can apply to them, and are projected by the rules onto the derivation under the combinatory projection principle (5), thereby imposing more or less rigid word-order on the language in question.

It is important at this point to note that $S\backslash NP$ and $S/NP$, the types of the tensed intransitive, are not yet syntactically typable, All verbs so far are functions into $VP_{e\to t}$, corresponding to the Minimalist vP. The stem types are then mapped onto further categories corresponding to Minimalist functional projections, in English either morphologically, as with tense, or in syntactic derivation, by auxiliary verbs, as with progressive and perfect aspect.

For example, English passive morphology and Dyirbal antipassive morphology map accusative and absolutive transitive VPs onto intransitive VP. Thus for English

(21) -en := $VP_{pssv}\$\backslash\backslash VP\$ : \lambda p\lambda x\ldots\lambda y.p\ldots x\,sk^{(x)}_{\lambda y.likely(p\,x\ldots y)}$

(where \$ schematizes over subcategorizations)

Similarly, past- and present- participial morphology turns English VP stems into participial forms. For example:

(22) -ing := $VP_{prog}\$\backslash\backslash VP\$ : \lambda p\ \ldots\ \lambda v\lambda r.ongoing\,(p\ldots)$

(23) -en := $VP_{pppl}\$\backslash\backslash VP\$ : \lambda p\ \ldots.consequent\,(p\ldots)$

Finally, either tense-morphology or a lexical rule maps infinitival verb stems $VP\$$ onto corresponding tensed forms of the same semantic type requiring a subject—for English, $S\backslash NP_{agr}\$$; for Welsh, $S\$/NP_{agr}$.

(24) -ed := $S\backslash NP_{agr}\$\backslash\backslash VP\$ : \lambda p\ \ldots.past(p\ldots)$

(Thus, morphological operators may either increase or decrease valency.)

The adjunct types, are also defined at this stage, differing only from the various aspectual phrase governors in being endotypic, or type-preserving, rather than exotypic or type-changing. Thus we have the following categories for "seems" and "quickly" of type$(e\to t)\to(e\to t)$

(25) a. seems := $(S\backslash NP)/VP : \lambda p\lambda y.seemingly\,p\,x$
b. quickly := $(S\backslash NP)/(S\backslash NP) : \lambda p\lambda y.manner\,(p\,y)\,quick$

Such adjuncts may be similarly morpholexically derived,

(26) -ly := $(VP\mid VP)\backslash\backslash(N\mid N) : \lambda a\lambda p\lambda y.manner\,(p\,y)\,a$

This third level of the lexical type system corresponds to the minimalist levels of "head-movement" (Roberts, 2001), "scrambling" (Ross, 1967), and "sideward move-

ment" (Nunes, 2001). The Head Movement Constraint (Travis, 1984) follows like other locality/minimality conditions from the fact that these relations are defined lexically.

A fourth stage of morpho-lexical derivation follows, in which case morphology (or "structurally" disambiguated underspecification) maps argument categories *NP*, *VP*, etc. and their heads or governing categories $NP/N$, $VP/NP$, etc. into order-preserving functors categories type-raised over all and only the lexical (tensed, nominal, participial, passive etc.) functor types that take them as arguments, such as $S/(S\backslash NP)$, $(S\backslash(S/NP))/N$.

In a final fifth stage of lexical derivation, the order-preserving cased categories that are specifically raised over $S\$$ are mapped to non-order-preserving lexical governors raised over tensed lexical types notably including $S/NP$, such as wh-words and *tough*-predicates, which differ from cased arguments in changing the type of their result to $N\backslash N$, $S\backslash NP$. In the case of topicalizers and *Wh*-question words, the changed category is a root construction $S_{top}$ or $S_{whq}$ for which no stem subcategorizes.

These levels feed each other sequentially in generating the lexicon. For example, the pied-piping *wh*-category of example (82) is repeated here:

(27) which :=$((N\backslash N)/(S/NP))\backslash(((S\backslash(S/NP))/NP))$

This category is derived for the lexicon as follows. $e \rightarrow e$ is a level 1 type, so $NP/NP$ is a level 3 syntactic type, so $(S\backslash(S/NP))/NP$ is a level 4 cased type, so $(S\backslash(S/NP))\backslash((S\backslash(S/NP))/NP)$ is a cased type, so $((N\backslash N)/(S/NP))\backslash((S\backslash(S/NP))/NP)$ is a level 5 *wh*-category.

All and only the Combinatory **B** and **S** Rules allowed under the CPP are then necessary to project the language specific lexical categories onto sentence derivations.

These various levels of lexicalization are reminiscent of Emonds's 1976 "structure-preserving", "local", and "root" transformations, and the related "levels" of movement phenomena of Williams, 2003 and Abels (2008), and the "tiers" of Culicover and Jackendoff, 2005. In particular, level 2 above corresponds to Abels' "A-mvt", level 3 to his "scrambling", and level 5 to his "*wh*-mvt" and "topicalization"

levels. (Level 4 of case assignment via type-raising is not recognized by Williams or Abels but may be related to Culicover and Jackendoff's "case tier".)[14]

The claim is that all and only the degrees of freedom this type system allows in the lexicon are observed in the languages of the world.

For example, we have noted that the type system allows the following categories for raising verbs:

14. It is interesting to speculate where this layered structure of levels stems from. Like all substantive universals, one plausible origin is the nature of the underlying cognitive representation. For example, Steedman (2002) suggests that the origin of type-raising/case lies in a represntation of objects in terms of the actions which they allow—in the terms of Gibson (1977), their "affordances"—a representation much of which which we share with other tool-using animals, notably chimpanzees.

(28) a. seems := $(S\backslash NP)/VP_{to} : \lambda p \lambda y.seemingly(py)$
b. seems := $(S\backslash NP_{xpl})/S : \lambda p \lambda s.seemingly\, s$

This allows the observed English raising constructions

(29) a. John seems to be certain to leave.
b. It seems (that) John is certain to leave.

But we cannot write categories that would allow "super-raising", as in (30b), whose exclusion motivated the introduction of the Minimal Link Condition (MLC) (cf. Chomsky 1995b:82,(131b)). First, there are no wrapping or commuting combinatory rules in CCG that would allow *seems* to reach across *(that) it is certain* to directly control *to leave*. To allow (30b) we would instead need something like the lexical category (30a) for "seemingly", which throws away the optional complementizer and the expletive, and applies to the non-adjacent finite predicate *is certain to leave* with the intended interpretation:

(30) a. *seems := $(((S\backslash NP_{3s})/(S_{fin}\backslash NP_{3s}))/NP_{xpl})/(S'/S)$
$: \lambda q \lambda x \lambda p \lambda y.seemingly(py)$
b. *John seems that it is certain to leave.

However, under the earlier definitions of the lexical type system and tense, tensed $S\backslash NP$ is not a possible argument for "seemingly" or any other verb, as opposed to adjuncts like "seemingly" (cf. Hornstein, 2009:168n25, who attributes this fact to the Principle of Greed, cf. Boeckx et al., 2010:ch.4):[15]

(31) a. *seems := $(S\backslash NP_{3s})/(S_{fin}\backslash NP) : \lambda p \lambda y.seemingly(py)$
b. *John seems is certain to leave.
c. John seemingly is certain to leave.

In fact, we have seen that the only categories that can subcategorize for tensed predicates are type-raised arguments, adjuncts, and the "*wh*-movers", such as topicalized and relativized arguments (which are in effect themselves type-raised).

Nor, as in the related framework of Cormack and Smith (2004), is it possible to write "backward control" categories, of the kind required by the analyses of Monahan (2003) and Polinsky (2012)—see Boeckx et al., 2010:106-114 for discussion—since control is defined as a relation between coarguments of a single verb.

Nevertheless, the following lexical entries are both licensed, despite the fact that the second involves a controller *y* that is not the closest candidate *x*, in apparent contradiction to the Minimal Link Condition (MLC) of Chomsky (1995b):

(32) a. persuades := $((S\backslash NP)/VP_{to})/NP : \lambda x \lambda p \lambda y.persuades(px)xy$
b. promises := $((S\backslash NP)/VP_{to})/NP : \lambda x \lambda p \lambda y.promises(py)xy$

15. Again the reason is ultimately semantic-conceptual.

The problem of how to limit movement so that it is bounded, excluding super-raising (31b) without at the same time excluding (32b), has been a problem since Rosenbaum, 1967 (cf. Hornstein, 2009:163-164), and was a major motivation for the PRO analysis of control, in which the binding of controlled elements is accomplished by extra-grammatical meaning postulates, rather than by movement. In the present framework, these constraints all follow from the fact that these constructions are lexicalized, rendering the MLC redundant.

Thus, the differing levels or domains of these authors' varieties of movement correspond to different levels of the morpholexicon: the stem which determines raising and control; the local operators such as tense, mood, aspect, and voice; the domain of case—that is, morpholexical order-preserving type-raising (including underspecification or "structurally" determined case), which determines scrambling, including so-called "long-range" scrambling; the domain of morpholexical type-changing higher-order categories such as relativizers, which determine the projection of lexical subcategorization onto unbounded dependencies by syntactic derivation. (Some of the latter, such as focalizing and topicalizing categories, limit the scope of the unbounded dependencies concerned to the root clause.)

However, none of these morpholexically-specified domains involves a distinct level of representation in the standard sense of the term. Each of them pairs an interpretation at the level of logical form—the only representational level that is countenanced in CCG—with a syntactic categorial type, which is the sole determinant of possible syntactic projections via the combinatory syntactic rules.

Up to this point, we have talked of the lexicon as if all forms related to a single stem are exhaustively listed. Given the degree of idiosyncrasy in the English lexicon, this is not unreasonable. Even in languages with highly productive morphology like Latin and Finnish, it may well be the most effficent way to run the processor. Nevertheless, to the extent that there are regularities across forms related to the same stem, it is useful to also have lexical rules rules expressing those regularities, so that once one form for a previously unseem stem is encountered, all the other forms can be predicted. However, whether such rules are used offline to compile those forms out in the lexicon in all their forms ("lexical redundancy rules"), or are used actively online ("lexical rules") is similarly a question of efficiency in implementation and emprical prediction for the psychologist. For purposes of the theory of grammar, all of these options are equivalent, so we will continue to ignore them here.

## 11.7 Envoi

Once an explanatory theory of grammar is achieved, the theory of linguistic competence is complete, and the purely theoretical side of the research program defined in *Syntactic Structures* is concluded.

However, an explanatory theory of grammar still does not constitute a complete the-

ory of Language. To explain how a child acquires that grammar requires a theory of how they search the space of possibilities allowed by that theory, and of the evidence that guides this search. This process in turn presupposes an account of linguistic performance or use. Finally one must ask how evolution could come up with that resource in what seems to have been a very short space of time indeed—at most a few million years.

These further questions of the space of possible grammars allowed by CCG and the nature of the processing mechanism, their use in a model of language acquisition, and their possible origins in biological evolution. all concern performance mechanims, rather than the competence grammar that is the focus of this book. They have accordingly been relegated to a series of brief appendices, which the purely linguistically inclined reader may prefer to skip.

**Exercise** : The system of lexical types outlined above just applies to the content words, and omits function words like conjunctions, negation, and the quantifier determiners. How could they be brought into the system?