The Combinatory Morphemic Lexicon

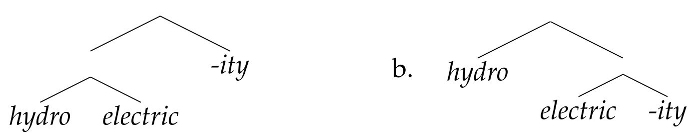
Cem Bozsahin*
Middle East Technical University

Grammars that expect words from the lexicon may be at odds with the transparent projection of syntactic and semantic scope relations of smaller units. We propose a morphosyntactic framework based on Combinatory Categorial Grammar that provides flexible constituency, flexible category consistency, and lexical projection of morphosyntactic properties and attachment to grammar in order to establish a morphemic grammar-lexicon. These mechanisms provide enough expressive power in the lexicon to formulate semantically transparent specifications without the necessity to confine structure forming to words and phrases. For instance, bound morphemes as lexical items can have phrasal scope or word scope, independent of their attachment characteristics but consistent with their semantics. The controls can be attuned in the lexicon to language-particular properties. The result is a transparent interface of inflectional morphology, syntax, and semantics. We present a computational system and show the application of the framework to English and Turkish.

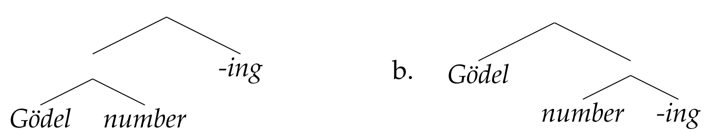
1. Introduction

The study presented in this article is concerned with the integrated representation and processing of inflectional morphology, syntax, and semantics in a unified grammar architecture. An important issue in such integration is mismatches in morphological, syntactic, and semantic bracketings. The problem was first noted in derivational morphology. Williams (1981) provided examples from English; the semantic bracketings in (1a–2a) are in conflict with the morphological bracketings in (1b–2b).

(1) a. b.



(2) a. b.



If the problem were confined to derivational morphology, we could avoid it by making derivational morphology part of the lexicon that does not interact with grammar. But this is not the case. Mismatches in morphosyntactic and semantic bracketing

* Computer Engineering and Cognitive Science, Middle East Technical University, 06531 Ankara, Turkey. E-mail: bozsahin@metu.edu.tr.

© 2002 Association for Computational Linguistics

also abound. This article addresses such problems and their resolution in a computational system.[1]

Müller (1999, page 401) exemplifies the scope problem in German prefixes. (3a) is in conflict with the bracketing required for the semantics of the conjunct (3b).

(3) a. *Wenn [Ihr Lust] und [noch nichts anderes vor-]habt,*
if you pleasure and yet nothing else intend

können wir sie ja vom Flughafen abholen
can we them PARTICLE from.the airport pick up
'If you feel like it and have nothing else planned, we can pick them up at the airport.'

b. *Ihr Lust habt UND noch nichts anderes vorhabt*

Similar problems can be observed in Turkish inflectional suffixes. In the coordination of tensed clauses, the tense attaches to the verb of the rightmost conjunct (4a) but applies to all conjuncts (4b). Delayed affixation appears to apply to all nominal inflections (4c–e).

(4) a. *Zorunlu deprem sigortası [yürürlüğe girmiş] ama*
mandatory earthquake insurance effect enter-ASP but

[tam anlamıyla uygulanamamış]-tı
exactly apply-NEG-ASP-TENSE
'Mandatory earthquake insurance had gone into effect, but it had not been enforced properly.'

b. *yürürlüğe girmiş-ti ama tam anlamıyla uygulanamamış-tı*

c. *Adam-ın [araba ve ev]-i*
man-GEN car and house-POSS
'the man's house and car'

d. *Araba-yı [adam ve çocuk]-lar-a göster-di-m*
Car-ACC man and child-PLU-DAT show-TENSE-PERS1
'(I) showed the car to the men and the children.'

e. *Araba-yı sen-in [dost ve tanıdık]-lar-ın-a göster-di-m*
Car-ACC you-GEN friend and acq.-PLU-POSS-DAT showed
'(I) showed the car to the your friends and acquaintances.'

1 Our use of the term **morphosyntax** needs some clarification. Some authors, (e.g., Jackendoff 1997), take it to mean the syntax of words, in contrast to the syntax of phrases. By morphosyntax we mean those aspects of morphology and syntax that collectively contribute to grammatical meaning composition. This is more in line with the inflectional-morphology-is-syntax view. In this respect, we will not address problems related to derivational morphology; its semantics is notoriously noncompositional and does not interact with grammatical meaning. Moreover, without a semantically powerful lexicon such as Pustejovsky's (1991), even the most productive fragment of derivational morphology is hard to deal with (Sehitoglu and Bozsahin 1999).

Phrasal scope of inflection can be seen in subordination and relativization as well. In (5a), the entire nominalized clause marked with the accusative case is the object of *want*. In (5b), the relative participle applies to the relative clause, which lacks an object. The object's case is governed by the subordinate verb, whose case requirements might differ from that of the matrix verb (5c). As we show later in this section, the coindexing mechanisms in word-based unification accounts of unbounded extraction face a conflict between the local and the nonlocal behavior of the relativized noun, mainly due to applying the relative participle *-diğ-i* to the verbal stem *ver* rather than the entire relative clause. A lexical entry for *-diğ-i* would resolve the conflict and capture the fact that it applies to nonsubjects uniformly.

(5) a. *Can* [*Ayşe'nin kitab-ı* *oku-ma-sı*]-*nı* *iste-di*
C.NOM A.-GEN book-ACC read-INF-AGR-ACC want-TENSE
'Can wanted Ayşe to read the book.'
lit. 'Can wanted Ayşe's-reading-the-book.'

b. *Ben* [*Mehmet'in çocuğ-a/*-u* *ver*]-*diğ-i* *kitab-ı* *oku-du-m*
I.NOM M-GEN child-DAT/*ACC give-REL.OP book-ACC read-TENSE-PERS1
'I read the book that Mehmet gave to the child.'

c. *Ben* [*Mehmet'in kitab-ı* *ver*]-*diğ-i* *çocuğ-u/*-a* *gör-dü-m*
I.NOM M-GEN book-ACC give-REL.OP child-ACC/*DAT see-TENSE-PERS1
'I saw the child to whom Mehmet gave the book.'

The morphological/phrasal scope conflict of affixes is not particular to morphologically rich languages. Semantic composition of affixes in morphologically simpler languages poses problems with word (narrow) scope of inflections. For instance, *fake trucks* needs the semantics (plu(fake truck)), which corresponds to the surface bracketing [*fake truck*]-*s*, because it denotes the nonempty nonsingleton sets of things that are not trucks but fake trucks (Carpenter 1997). *Four trucks*, on the other hand, has the semantics (four(plu truck)), which corresponds to *four* [*truck*]-*s*, because it denotes the subset of nonempty nonsingleton sets of trucks with four members.

The status of inflectional morphology among theories of grammar is far from settled, but, starting with Chomsky (1970), there seems to be an agreement that derivational morphology is internal to the lexicon. Lexical Functional Grammar (LFG) (Bresnan 1995) and earlier Government and Binding (GB) proposals e.g. (Anderson 1982) consider inflectional morphology to be part of syntax, but it has been delegated to the lexicon in Head-Driven Phase Structure Grammar (HPSG) (Pollard and Sag 1994, page 35) and in the Minimalist Program (Chomsky 1995, page 195). The representational status of the morpheme is even less clear. Parallel developments in computational studies of HPSG propose lexical rules to model inflectional morphology (Carpenter and Penn 1994). Computational models of LFG (Tomita 1988) and GB (Johnson 1988; Fong 1991), on the other hand, have been noncommittal regarding inflectional morphology. Finally, morphosyntactic aspects have always been a concern in Categorial Grammar (CG) (e.g., Bach 1983; Carpenter 1992; Dowty 1979; Heylen 1997; Hoeksema 1985; Karttunen 1989; Moortgat 1988b; Whitelock 1988), but the issues of constraining the morphosyntactic derivations and resolving the apparent mismatches have been relatively untouched in computational studies.

We briefly look at Phrase Structure Grammars (PSGs), HPSG, and Multimodal CGs (MCGs) to see how word-based alternatives for morphosyntax would deal with the issues raised so far. For convenience, we call a grammar that expects words from the lexicon a **lexemic** grammar and a grammar that expects morphemes a **morphemic** grammar. A lexemic PSG provides a lexical interface for inflected words (X_0s) such that a regular grammar subcomponent handles lexical insertion at X_0.[2] In (4d), the right conjunct *çocuk-lar-a* is analyzed as $N_0 \rightarrow$ çocuk-PLU-DAT (or $N_0 \rightarrow N_{0'}$ -DAT, $N_{0'} \rightarrow N_{0''}$ -PLU, $N_{0''} \rightarrow$ Stem, as a regular grammar). Assuming a syncategorematic coordination schema, that is, $X \rightarrow X \textit{ and } X$, the N_0 in the left and right conjuncts of this example would not be of the same type. Revising the coordination schema such that only the root features coordinate would not be a solution either. In (4e), the relation of possession that is marked on the right conjunct must be carried over to the left conjunct as well. What is required for these examples is that the *syntactic* constituent X in the schema be analyzed as X-PLU(-POSS)-DAT, after $N_0 \textit{ and } N_0$ coordination.

What we need then is not a lexemic but a morphemic organization in which bracketing of free and bound morphemes is regulated in syntax. The lexicon, of course, must now supply the ingredients of a morphosyntactic calculus. This leads to a theory in which semantic composition parallels morphosyntactic combination by virtue of bound morphemes' being able to pick their domains just like words (above X_0, if needed). A comparison of English and Turkish in this regard is noteworthy. The English relative pronouns *that/whom* and the Turkish relative participle *-diğ-i* would have exactly the same semantics when the latter is granted a representational status in the lexicon (see Section 6).

Furthermore, rule-based PSGs project a rigid notion of surface constituency. Steedman (2000) argued, however, that syntactic processes such as identical element deletion under coordination call for flexible constituency, such as SO (subject-object) in the SVO & SO gapping pattern of English and SV (subject-verb) constituency in the OSV & SV pattern of Turkish. Nontraditional constituents are also needed in specifying semantically transparent constituency of words, affixes, clitics, and phrases.

Constraint-based PSGs such as HPSG appeal to coindexation and feature passing via unification, rather than movement, to deal with such processes. HPSG also makes the commitment that inflectional morphology is internal to the lexicon, handled either by lexical rules (Pollard and Sag 1994) or by lexical inheritance (Miller and Sag 1997). We look at (5c) to highlight a problem with the stem-and-inflections view. As words enter syntax fully inflected, the sign of the verb *ver-diğ-i* in the relative clause (5c) would be as in (6a), in which the SUBCAT list of the verb stem is, as specified in the lexical entry for *ver*, unsaturated. The participle adds coindexation in MOD$|\cdots|$INDEX. The HPSG analysis of this example would be as in Figure 1. Although passing the agreement features of the head separately (Sehitoglu 1996) solves the case problem alluded to in (5c), however, structure sharing of the NP_{dat} with the SLASH, INDEX, and CONTENT features of *ver-diğ-i* is needed for semantics (GIVEE), but this conflicts with the head features of the topmost NP_{acc} in the tree. The relative participle as a lexical entry (e.g., (6b)) would resolve the problem with subcategorization because its SUBCAT list is empty (like the relative pronoun *that* in English), hence there would be no indirect dependence of the nonlocal SLASH feature and the local SUBCAT feature via semantics (CONTENT). Such morphemic alternatives are not considered in HPSG, however, and require a significant revision in the theory. Furthermore, HPSG's lexical

2 But see Creider, Hankamer, and Wood (1995), which argues that the morphotactics of human languages is not regular but linear context free.

assignment for trace introduces phonologically null elements into the lexicon, which, as we show later, is not necessary.

(6) a. ver-diğ-i :=

$$\begin{bmatrix} \text{LOCAL} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR} & \begin{bmatrix} \text{PERSON} & \textit{third} \\ \text{NUMBER} & \textit{sing} \end{bmatrix} \\ \text{CASE} & \textit{dat} \end{bmatrix} \\ \text{SUBCAT} & \langle \boxed{3}\ \text{NP[gen]}, \boxed{2}\ \text{NP[acc]}, \boxed{1}\ \text{NP[dat]} \rangle \\ \text{MOD} \mid \text{MODSYN} \mid \text{LOCAL} \mid \text{CONT} \mid \text{INDEX} & \boxed{1} \end{bmatrix} \\ \text{CONTENT} & \begin{bmatrix} \text{RELN} & \textit{give} \\ \text{GIVER} & \boxed{3} \\ \text{GIVEE} & \boxed{1} \\ \text{GIFT} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{NONLOCAL} \mid \text{TO-BIND} \mid \text{SLASH} & \left\{ \boxed{1} \right\} \end{bmatrix}$$

b. -diğ-i :=

$$\begin{bmatrix} \text{LOCAL} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \begin{bmatrix} \text{acc or dat} \end{bmatrix} \\ \text{SUBCAT} & \langle\rangle \end{bmatrix} \\ \text{CONTENT} & \textit{npro} \begin{bmatrix} \text{INDEX} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{NONLOCAL} \mid \text{INHER} \mid \text{SLASH} & \left\{ \boxed{1} \right\} \end{bmatrix}$$

MCGs (Hepple 1990a; Morrill 1994; Moortgat and Oehrle 1994) allow different modes of combination in the grammar. In addition to binary modes such as wrapping and commutative operations, unary modalities provide finer control over the categories. Heylen (1997, 1999) uses unary modalities as a way of regulating morphosyntactic features such as case, number, and person for economy in lexical assignments. For instance, *Frau* has the category $\Box_{case}\Box_{fem}\Box_{sg}\Box_{3p}\Box_{decl}N$, which underspecifies it for case and declension. Underspecification is dealt with in the grammar using inclusion postulates (e.g., (7)). The interaction of different modalities is regulated by distribution postulates.

(7) $$\frac{\Box_{case}\Gamma \vdash X}{\Box_{nom}\Gamma \vdash X} \qquad \frac{\Box_{case}\Gamma \vdash X}{\Box_{acc}\Gamma \vdash X}$$

Lexical assignments to inflected words carry unary modalities: *boys* has the type $\Box_{pl}N$, in contrast to $\Box_{sg}N$ for *boy*. Although such regulation of inflectional features successfully mediates, for example, subject-verb agreement or NP-internal case agreement (as in German), it is essentially word-based, because type assignments are to inflected forms; morphemes do not carry types. This reliance on word types necessitates a lexical rule–based approach to some morphosyntactic processes that create indefinitely long words, such as *ki*-relativization in Turkish (see Section 6.5). But lexical rules for such processes risk nontermination (Sehitoglu and Bozsahin 1999). Our main point of departure from MCG accounts is the morphemic versus lexemic nature of the lexicon: The morphosyntactic and attachment modalities originate from the lexicon; they are not properties of the grammar (we elaborate more on this later). This paves the way to the morphemic lexicon by licensing type assignments to units smaller than words.

Besides problems with lexical rules, the automata-theoretic power of MCGs is problematic: Unrestricted use of structural modalities and postulates leads to Turing completeness (Carpenter 1999). Indeed, one of the identifiable fragments of Mul-

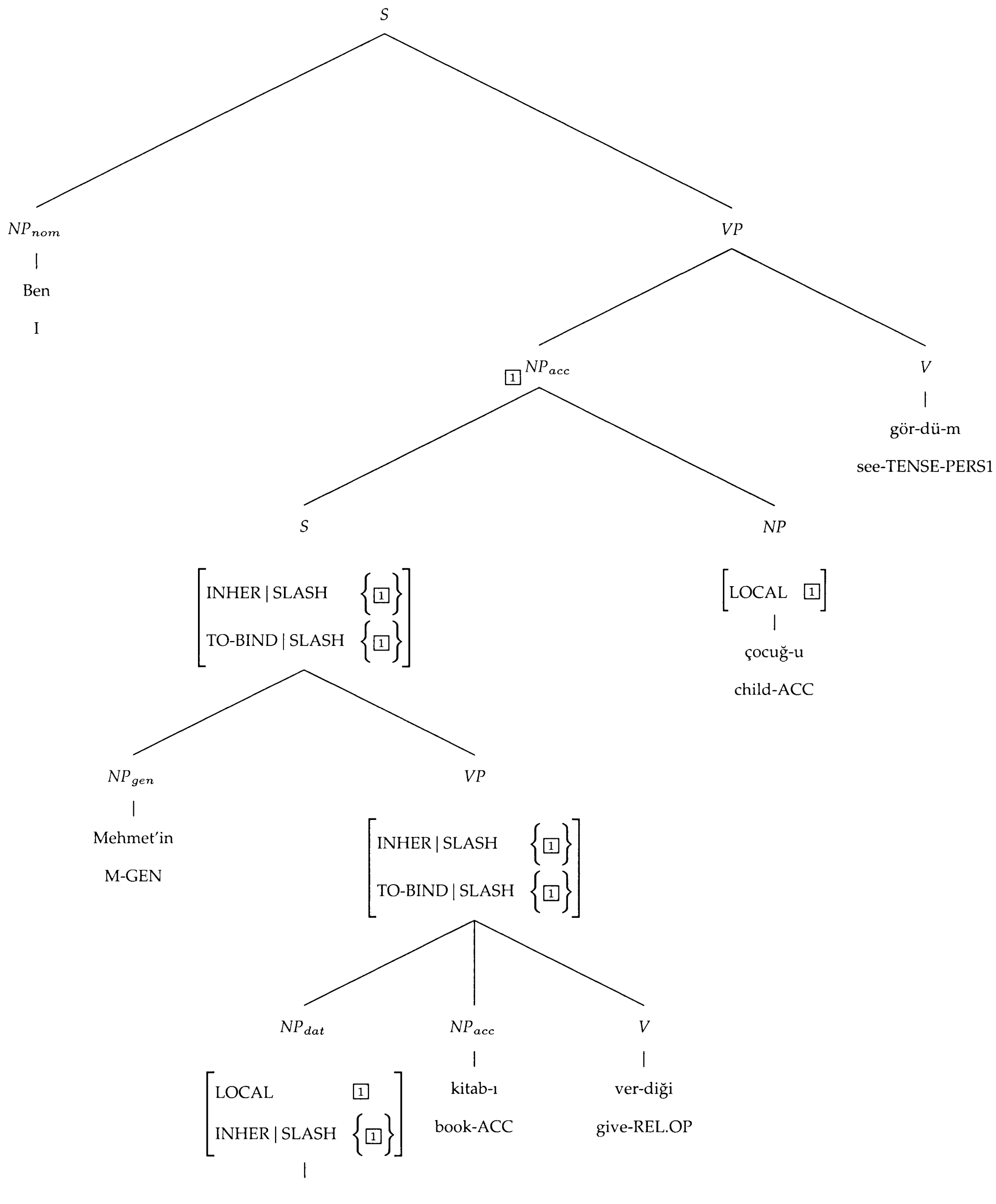


Figure 1
HPSG analysis of (5c).

timodal languages that is computationally tractable is Combinatory Categorial languages (Kruijff and Baldridge 2000), which we adopt as the basis for the framework presented here. We propose a morphosyntactic Combinatory Categorial Grammar (CCG) in which the grammar and the morphemic lexicon refer to morphosyntactic types rather than syntactic types. We first introduce the syntactic CCG in Section 2. Morphosyntactic CCG is described in Section 3. In Section 4, we look at the computational aspects of the framework. We then show its realization for some aspects of English (Section 5) and Turkish (Section 6).

2. Syntactic Types

CG is a theory of grammar in which the form-meaning relation is conceived as a transparent correspondence between the surface-syntactic and semantic combinatorics (Jacobson 1996). A CCG sign can be represented as a triplet $\pi - \sigma : \mu$, where π is the prosodic element, σ is its syntactic type, and μ its semantic type. For instance, the lexical assignment for *read* is (8).[3]

(8) read := $read - (S\backslash NP)/NP : \lambda x.\lambda y.\mathsf{read}\, xy$

Definition (Syntactic Types)

- The set of basic syntactic categories: $\mathcal{A}_s = \{N, NP, S, S_{-t}, S_{+t}\}$
- The set of complex syntactic categories: $\mathcal{B}_s$
  - — $\mathcal{A}_s \subseteq \mathcal{B}_s$
  - — If $X \in \mathcal{B}_s$ and $Y \in \mathcal{B}_s$, then $X\backslash Y$ and $X/Y \in \mathcal{B}_s$

The classical Ajdukiewicz/Bar-Hillel (AB) CG is weakly equivalent to Context-Free Grammars (Bar-Hillel, Gaifman, and Shamir 1960). It has function application rules, defined originally in a nondirectional fashion. The directional variants and their associated semantics are as follows:

(9) Forward Application (>):[4] $X/Y : f \quad Y : a \quad \Rightarrow \quad X : fa$
Backward Application (<): $Y : a \quad X\backslash Y : f \quad \Rightarrow \quad X : fa$

CCG (Steedman 1985, 1987, 1988; Szabolcsi 1983, 1987) is an extended version of AB that includes function composition (10), substitution, and type raising (11). These extensions make CCGs mildly context sensitive.

(10) Forward Composition (>B): $X/Y : f \quad Y/Z : g \quad \Rightarrow \quad X/Z : \lambda x.f(gx)$
Backward Composition (<B): $Y\backslash Z : g \quad X\backslash Y : f \quad \Rightarrow \quad X\backslash Z : \lambda x.f(gx)$

(11) Forward Type Raising (>T):[5] $X : a \quad \Rightarrow \quad T/(T\backslash X) : \lambda f.f[a]$
Backward Type Raising (<T): $X : a \quad \Rightarrow \quad T\backslash(T/X) : \lambda f.f[a]$

Type raising is an order-preserving operation. For instance, Lambek's (1958) category $S/(S\backslash NP)$ is a positional encoding of the grammatical subject as a function

3 We take π to be the surface string for simplicity. We use the "result-first" convention for CG. For instance, transitive verbs of English are written as $(S\backslash NP)/NP$, which translates to $(NP\backslash S)/NP$ in the "result-on-top" convention.

4 We omit the prosodic element for ease of exposition. For instance, the complete definition of forward application is $s_1 - X/Y : f \quad s_2 - Y : a \quad \Rightarrow \quad s_1 \bullet s_2 - X : fa$, where $\bullet$ is prosodic combination and fa is the application of f to a. The $\bullet$ will play a crucial role in the lexicalization of attachment later on.

5 The lambda term $f[a]$ denotes internal one-step β-reduction of f on a. In parsing, we achieve the same effect by partial execution (Pereira and Shieber 1987). $\lambda f.f[a]$ is encoded as `(a^F)^F` in Prolog, where `^` is lambda abstraction. We opted for the explicit $f[a]$ notation mainly for ease of exposition (cf. the semantics of raising verbs, relative participles, etc. in Section 6). Moreover, as Pereira and Shieber noted, `(a^F)^F` is not a lambda term in the strict sense because `a` is not a variable.

looking for a VP ($= S\backslash NP$) to the right to become S. The reversal of directionality such as topicalization (e.g., *This book, I recommend*) requires another schema. The reversal is with respect to the position of the verb, which we shall call **contraposition** and formulate as in (12).[6] (<XP) is leftward extraction of a right constituent, and (>XP) is rightward extraction of a left constituent, both of which are marked constructions. Directionally insensitive types such as $T|(T|X)$ cause the collapse of directionality in surface grammar (Moortgat 1988a).

(12) Leftward Contraposition (<XP): $X{:}a \Rightarrow S_{+t}/(S/X){:}\lambda f.f[a]$, $S_{+t}/(S_{+t}/X){:}\lambda f.f[a]$

Rightward Contraposition (>XP): $X{:}a \Rightarrow S_{-t}\backslash(S\backslash X){:}\lambda f.f[a]$, $S_{-t}\backslash(S_{-t}\backslash X){:}\lambda f.f[a]$

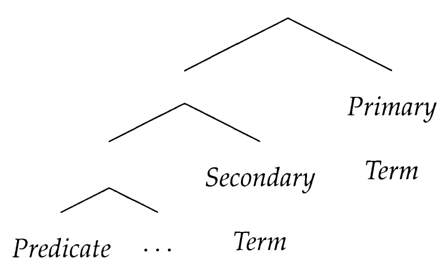
The semantics of contraposition depends on discourse properties as well. We leave this issue aside by (*1*) noting that it is related to type raising in changing the function-argument relation and (*2*) categorizing the sentence as S_{+t} (topicalized) or S_{-t} (detopicalized), which are not discourse equivalent to S. Syntactic characterization as such also helps a discourse component do its work on syntactic derivations.

CCG's notion of interpretation is represented in the Predicate-Argument Structure (PAS). Its organization is crucial for our purposes, since the bracketing in the PAS is the arbitrator for reconciling the bracketings in morphology and syntax via proper lexical type assignments. It is the sole level of representation in CCG (Steedman 1996, page 89).[7] It is the level at which the conditions on objects of interpretation, such as binding and control, are formulated. For instance, Steedman (1996) defines c-command and binding conditions A, B, and C over the PAS. The PAS also reflects the obliqueness order of the arguments:

Predicate ... Tertiary-Term Secondary-Term Primary-Term

Assuming left associativity for juxtaposition, this representation yields the bracketing in (13) for the PAS. Having the primary argument as the outermost term is motivated by the observations on binding asymmetries between subjects and complements in many languages (e.g., **Himself saw John*, **heself*).

(13)



3. Morphosyntactic Types

A syntactic type such as N does not discriminate morphosyntactically. A finer distinction can be made as singular nouns, plural nouns, case-marked nouns, etc. For

6 In fact, topicalization of nonperipheral arguments (*This book, I would give to Mary*) requires that (12) be finitely schematized over valencies, such as S, S/NP, S/PP (Steedman 1985).

7 We will not elaborate on the theoretical consequences of having this level of representation; see, for instance, Dowty (1991) and Steedman (1996).

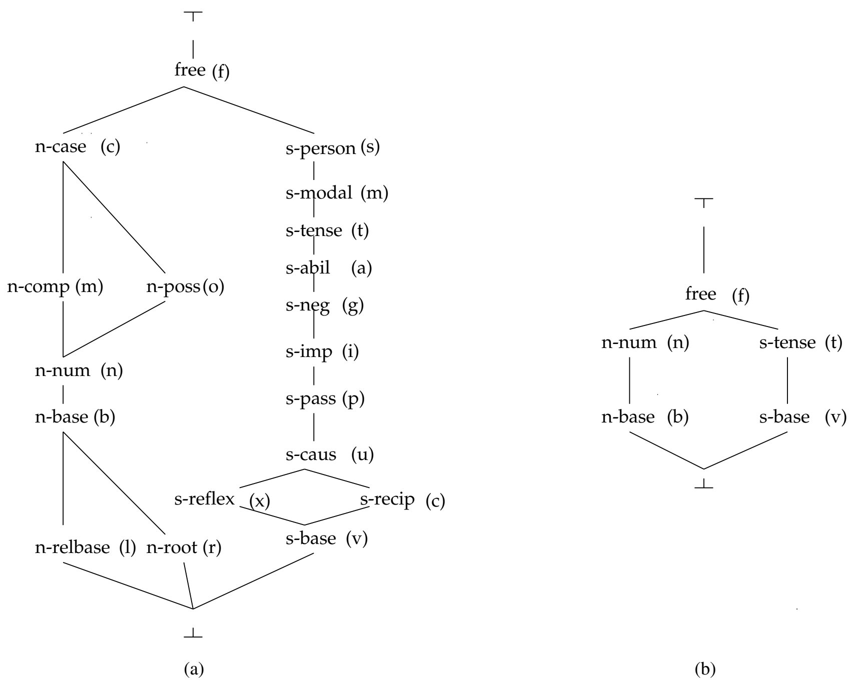


Figure 2
The lattice of diacritics for (a) Turkish and (b) English.

instance, the set of number-marked nouns can be represented as $\overset{n}{\bowtie}N$, where $\bowtie$ is a morphosyntactic modality ("equals") and n is a diacritic (for *n*umber). *Books* is of type $\overset{n}{\bowtie}N$, but *book* is not. The type for *books* can be obtained morphosyntactically by assigning *-s* (-PLU) the functor type $\overset{n}{\bowtie}N\backslash\overset{b}{\bowtie}N$, where b stands for *b*ase. A syntactic type such as $N\backslash N$ overgenerates.

Another modality, $\triangleleft$ ("up to and equals"), allows wider domains in morphosyntactic typing. For instance, $\overset{n}{\triangleleft}N$ represents the set of nouns marked on number or any other diacritic that is lower than number in a partial order (e.g., Figure 2). The inflectional paradigm of a language can be represented as a partial ordering using the modalities.[8] For instance, if the paradigm is Base-Number-Case, we have $\upsilon(\overset{b}{\triangleleft}N) \subseteq \upsilon(\overset{n}{\triangleleft}N) \subseteq \upsilon(\overset{c}{\triangleleft}N)$, where $\upsilon(\tau)$ is the valuation function from the morphosyntactic type τ to the set of strings that have the type τ. The $\bowtie$ modality is more strict than $\triangleleft$ to provide finer control; although $\upsilon(\overset{n}{\triangleleft}N) \subseteq \upsilon(\overset{c}{\triangleleft}N)$, $\upsilon(\overset{n}{\bowtie}N) \not\subseteq \upsilon(\overset{c}{\bowtie}N)$, because a noun can be number marked but not case marked or vice versa. Also, $\upsilon(\overset{i}{\bowtie}N) \subseteq \upsilon(\overset{i}{\triangleleft}N)$ for any diacritic i since, for instance, the set of nouns marked up to and including case includes case-marked, number-marked, and unmarked nouns.

The lattice consistency condition is imposed on the set of diacritics to ensure category unity.[9] In other words, the syntactic type X can be viewed as an abbreviation for the morphosyntactic type $\overset{\top}{\triangleleft}X$ where $\top$ is the universal upper bound. It is the

8 See Heylen (1997) on use of unary modalities for a similar purpose in lexemic MCG.
9 In a lattice L, $x \leq y$ (morphosyntactically, $x \triangleleft y$) is equivalent to the consistency properties $x \wedge y = x$ and $x \vee y = y$. We use the join operator for this check, thus it suffices to have a join semilattice.

most underspecified category of X which subsumes all morphosyntactically decorated versions of X. Figure 2 shows the lattice for English and Turkish.

Definition (Morphosyntactic Types)

- $\mathcal{D}$ = finite set of diacritics
- Join semilattice $L = (\mathcal{D}, \leq, =)$
- The set of basic morphosyntactic types: $\mathcal{A}_{ms}$.
  - — $\overset{i}{\triangleleft} X \in \mathcal{A}_{ms}$ and $\overset{i}{\bowtie} X \in \mathcal{A}_{ms}$ if $i \in \mathcal{D}$ and $X \in \mathcal{A}_s$ (see definition of syntactic types for $\mathcal{A}_s$)
  - — ($\bowtie$ corresponds to lattice condition =)
  - — ($\triangleleft$ corresponds to lattice condition $\leq$)
- The set of complex morphosyntactic types: $\mathcal{B}_{ms}$
  - — $\mathcal{A}_{ms} \subseteq \mathcal{B}_{ms}$
  - — If $X \in \mathcal{B}_{ms}$ and $Y \in \mathcal{B}_{ms}$, then $X\backslash Y$ and $X/Y \in \mathcal{B}_{ms}$

For instance, the infinitive marker *-ma* in (14a) can be lexically specified to look for untensed VPs—functions onto $\overset{a}{\triangleleft} S$—to yield a complex noun base (14b), which, as a consequence of nominalization (result type *N*), receives case to become an argument of the matrix verb. The adjective in *fake trucks* can be restricted to modify unmarked *N*s to get the bracketing [*fake truck*]*-s* (14c).

(14) a. *Mehmet* [[*kitab-ı* *oku*]-*ma*]-*yı* *istiyor*
M.NOM book-ACC read-INF-ACC wants
'Mehmet wants to read the book.'

b. -INF := *ma* – $\overset{b}{\triangleleft} N\backslash(\overset{a}{\triangleleft} S\backslash \overset{f}{\triangleleft} NP_{nom})\colon \lambda f.f$

c. fake := *fake* – $\overset{b}{\triangleleft} N/\overset{b}{\triangleleft} N\colon \lambda x.\mathsf{fake}\,x$

Different attachment characteristics of words, affixes, and clitics must be factored into the prosodic domain as a counterpart of refining the morphosyntactic description. In Montague Grammar, every syntactic rule is associated with a certain mode of attachment, and this tradition is followed in MCG; attachment types are related with the slash (e.g., $/_w$ for wrapping), which is a grammatical modality.[10] In the present framework, however, attachment is projected from the lexicon to the grammar as a prosodic property of the lexical items.[11] The grammar is unimodal in the sense that / and \ simply indicate the function-argument distinction in adjacent prosodic elements. The lexical projection of attachment further complements the notion of morphemic lexicon so that bound morphemes are no longer parasitic on words but have an independent

10 See Dowty (1996) and Steedman (1996) for a discussion of bringing nonconcatenative combination into grammar.

11 There is a precedent of associating attachment characteristics with the prosodic element rather than the slash in CG (Hoeksema and Janda 1988). In Hoeksema and Janda's notation, arguments can be constrained on phonological properties and attachment. For instance, the English article *a* has its NP/N category spelled out as </CX/N,NP,Pref>, indicating a consonantal first segment for the noun argument and concatenation to the left.

Table 1
Attachment properties of some Turkish morphemes.

uzun (long) :=	$\overset{s}{\circ}\ uzun\ -\ \overset{b}{\triangleleft} N/\overset{b}{\triangleleft} N$	*uzun* *yol* long road 'long road'
oku (read) :=	$\overset{s}{\circ}\ oku\ -\ \overset{v}{\triangleleft} S\backslash\overset{f}{\triangleleft} NP_{nom}\backslash\overset{f}{\triangleleft} NP_{acc}$	*adam* *kitab-ı* *oku-du* man book-ACC read-TENSE 'the man read the book.'
-EMPH :=	$\overset{c}{\circ}\ de\ -\ X\backslash X$	*Ben* *de* *yaz-ar-ım* I too write-TENSE-PERS 'I write too.'
-LOC :=	$\overset{a}{\circ}\ de\ -\ \overset{c}{\triangleleft} N\backslash\overset{o}{\triangleleft} N$	*Ben-de* *kalem* *var* I-LOC pen exist 'I have a pen.'

representational status of their own. We write $\overset{i}{\circ} s$ to denote the attachment modality i (*a*ffixation, *s*yntactic concatenation, *c*liticization) of the prosodic element s.

Table 1 shows some lexical assignments for Turkish (e.g., the sign $\overset{a}{\circ} s - X\backslash Y{:}\,\mu$ characterizes a suffix). The morphosyntactic calculus of CCG is defined with the addition of morphosyntactic types and attachment modalities as follows (similarly, for other combinatory rules):

(15) Forward Application (>):

$$\frac{\overset{i}{\circ} s_1 - X/\overset{\alpha_1}{\Box_1} Y{:}f \qquad\qquad \overset{j}{\circ} s_2 - \overset{\alpha_2}{\Box_2} Y{:}a}{\overset{k}{\circ}(s_1 \overset{k}{\bullet} s_2) - X{:}fa}\ >$$

if $\alpha_2 \Box_1 \alpha_1$ in lattice L, for: $\Box_1, \Box_2 \in \{\bowtie, \triangleleft\}$, $\alpha_1, \alpha_2 \in \mathcal{D}$ in L, $i, j, k \in \{a, s, c\}$, $\overset{i}{\circ}\ \overset{j}{\circ} \vdash_a \overset{k}{\circ}$

Forward Composition (>B):

$$\frac{\overset{i}{\circ} s_1 - X/\overset{\alpha_1}{\Box_1} Y{:}f \qquad\qquad \overset{j}{\circ} s_2 - \overset{\alpha_2}{\Box_2} Y/Z{:}\,g}{\overset{k}{\circ}(s_1 \overset{k}{\bullet} s_2) - X/Z{:}\,\lambda x.f(gx)}\ >\mathrm{B}$$

if $\alpha_2 \Box_1 \alpha_1$ in lattice L, for: $\Box_1, \Box_2 \in \{\bowtie, \triangleleft\}$, $\alpha_1, \alpha_2 \in \mathcal{D}$ in L, $i, j, k \in \{a, s, c\}$, $\overset{i}{\circ}\ \overset{j}{\circ} \vdash_a \overset{k}{\circ}$

The main functor's argument specification ($\Box_1$ of $\overset{\alpha_1}{\Box_1} Y$ in (15)) determines the lattice condition in derivations.[12] Hence the morphosyntactic decoration in lexical assignments propagates its lattice condition to grammar as in $\alpha_2 \Box_1 \alpha_1$ (cf. Heylen [1997], in which the grammar rule imposes a *fixed* partial order, e.g., X/Y combines with Z if

12 This coincides with Steedman's (1991b) observation that directionality of the main functor's slash is also a property of the same argument. The main functor is the one whose result type determines the overall result type (i.e., X/Y in (15)).

$Z \leq Y$). This is another prerequisite that must be fulfilled for the morphemic lexicon to project the lexical specification of scope.

The grammar is not fixed on the attachment modality either (unlike a lexemic grammar, which is fixed on combination of words). Hence another requirement is the propagation of attachment to grammar. This is facilitated by the lexical types $\overset{m}{\circ} s - \sigma : \mu$, where m is an attachment type. The attachment calculus $\overset{i}{\circ} \; \overset{j}{\circ} \vdash_a \overset{k}{\circ}$ in (15), which reads "attachment types i and j yield type k," relates attachment to prosodic combination in the grammar.[13] It can be attuned to language-particular properties.

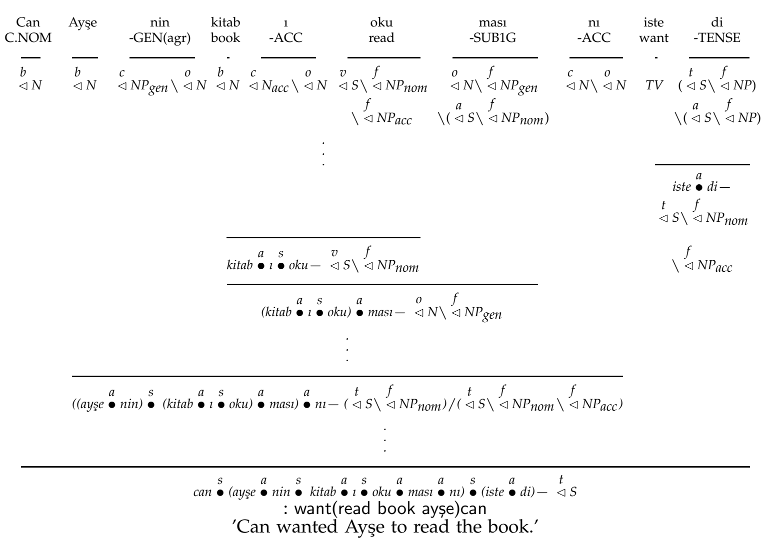
We can specify some prosodic properties of the attachment calculus for Turkish as follows ($\acute{x}$ indicates stress on the prosodic element x):

syntactic concatenation	$\acute{x} \overset{s}{\bullet} \acute{y}$	=	$\acute{x}\acute{y}$
affixation	$\acute{x} \overset{a}{\bullet} y$	=	$x\acute{y}$
cliticization	$\acute{x} \overset{c}{\bullet} y$	=	$\acute{x}\acute{y}$

4. Morpheme-Based Parsing

To contrast lexemic and morphemic processing, consider the Turkish example in (16a). We show some stages of the derivation to highlight prosodic combination (•) as well. Every item in the top row is a lexical entry. Allomorphs, such as that of tense, have the same category in the lexicon (16b). Vowel harmony, voicing, and other phonological restrictions are handled as constraints on the prosodic element. Constraint checking can be switched off during parsing to obtain purely morphosyntactic derivations.

(16) a.



b. -TENSE := $\overset{a}{\circ} dı|di|du|dü|tı|ti|tu|tü - (\overset{t}{\triangleleft} S \backslash \overset{f}{\triangleleft} NP) \backslash (\overset{a}{\triangleleft} S \backslash \overset{f}{\triangleleft} NP) : \lambda f.f$

13 Clearly, much more needs to be done to incorporate intonation into the system. The motive for attachment types is to provide the representational ingredients on behalf of the morphemic lexicon. As one reviewer noted, CCG formulation of the syntax-phonology interface moved from autonomous prosodic types (Steedman 1991a) to syntax-directed prosodic features (Steedman 2000b). The present proposal for attachment modality is computationally compatible with both accounts: Combinatory prosody can match prosodic types with morphosyntactic types. Prosodic features are associated with the basic categories of a syntactic type in the latter formulation, hence they become part of the featural inference that goes along with the matching of categories in the application of combinatory rules.

The lexicalization of attachment modality helps to determine the prosodic domain of postconditions. For instance, for Turkish, vowel harmony does not apply over word boundaries, which can be enforced by applying it when the modality is $\overset{a}{\circ}$ and $\overset{c}{\circ}$, but not $\overset{s}{\circ}$. Voicing applies to $\overset{a}{\circ}$ and $\overset{s}{\circ}$, but not to $\overset{c}{\circ}$.

The basic categories *N, NP, S,* $S_{+\mathrm{t}}$, and $S_{-\mathrm{t}}$ carry agreement features of fixed arity (e.g., tense and person for *S*, $S_{+\mathrm{t}}$, and $S_{-\mathrm{t}}$, and case, number, person, and gender for *N* and *NP*). Positional encoding of such information as in Pulman (1996) allows efficient term unification for the propagation of these features.[14] Term unification also handles the matching of complex categories in the CCG schema. For instance, $\overset{\alpha_1}{\Box_1}A/(\overset{\alpha_2}{\Box_2}B\backslash\overset{\alpha_3}{\Box_3}C)$ combines with $\overset{\beta_2}{\Box_4}B\backslash\overset{\beta_3}{\Box_5}C$ via ($>$) for $B, C \in \mathcal{A}_s$, if $\beta_2 \Box_2 \alpha_2$, $\beta_3 \Box_3 \alpha_3$ ($\Box_i \in \{\triangleleft, \bowtie\}$). Apart from the matching of syntactic types and agreement, unification does no linguistic work in this framework, in contrast to structure-sharing in HPSG and slash passing in Unification CG (Calder, Klein, and Zeevat 1988).

CCG is worst-case polynomially parsable (Vijay-Shanker and Weir 1993). This result depends on the finite schematization of type raising and bounded composition. Assuming a maximum valence of four in the lexicon (Steedman 2000a), composition (B^n) is bounded by $n \leq 3$. The refinement of the type raising schema (11) for finite schematization is shown in (17).

(17) a. Revised Forward Type Raising ($>$T): $NP\!:\!a \quad \Rightarrow \quad T/(T\backslash NP)\!:\!\lambda f.f[a]$
b. Revised Backward Type Raising ($<$T): $NP\!:\!a \quad \Rightarrow \quad T\backslash(T/NP)\!:\!\lambda f.f[a]$
$T \in \{S, S\backslash NP, S\backslash NP\backslash NP, S\backslash NP\backslash NP\backslash NP\}$.

The finite schematization of type raising suggests that it can be delegated to the lexicon, for example, by a lexical rule that value-raises all functions onto *NP* to their type-raised variety, such as NP/N to $(S/(S\backslash NP))/N$. But this move presupposes the presence of such functions in the lexicon, that is, a language with determiners. To be transparent with respect to the lexicon, we make type raising and other unary schema (contraposition) available in the grammar. Since both are finite schemas in the revised formulation, the complexity result of Vijay-Shanker and Weir still holds. Checking the lattice condition as in (15) incurs a constant factor with a finite lattice.

Type raising and composition cause the so-called spurious-ambiguity problem (Wittenburg 1987): Multiple analyses of semantically equivalent derivations are possible in parsing. This is shown to be desirable from the perspective of prosody; for example, different bracketings are needed to match intonational phrasing with syntactic structure (Steedman 1991). From the parsing perspective, the redundancy of analyses can be controlled by *(1)* grammar rewriting (Wittenburg 1987), *(2)* checking the chart for PAS equivalence (Karttunen 1989; Komagata 1997), *(3)* making the processor parsimonious on using long-distance compositions (Pareschi and Steedman 1987), or *(4)* parsing into normal forms (Eisner 1996; Hepple 1990b; Hepple and Morrill 1989; König 1989; Morrill 1999). We adopt Eisner's method, which eliminates chains of compositions in $O(1)$ time via tags in the grammar, before derivations are licensed. There is a switch that can be turned off during parsing to obtain all surface bracketings.

14 Mediating agreement via unification, type subsumption, or set-valued indeterminacy has important consequences on underspecification, the domain of agreement, and the notion of "like categories" in coordination (see Johnson and Bayer 1995; Dalrymple and Kaplan 2000; Wechsler and Zlatić 2000). Rather than providing an elaborate agreement system, we note that Pulman's techniques provide the mechanism for implementing agreement as atomic unification, subsumption hierarchies represented as lattices, or set-valued features. The categorial ingredient of phrase-internal agreement can be provided by endotypic functors when necessary (see Sections 5 and 6).

There is also a switch for checking the PAS equivalence, with the warning that the equivalence of two lambda expressions is undecidable.

The parser is an adaptation of the Cocke-Kasami-Younger (CKY) algorithm (Aho and Ullman 1972, page 315), modified to handle unary rules as well: In the kth iteration of the CKY algorithm to build constituents of length k, the unary rules apply to the CKY table entries $T[\alpha_i, \alpha_{i+k}], i = 0, 1, \ldots, n-k$; that is, k-length results of binary rules are input to potential unary constituents of length k. In practice, this allows, for instance, a nominalized clause to be type-raised after it is derived as a category of type N. The remaining combinatory schema is already in Chomsky Normal Form, as required by CKY. The finite schematization of CCG rules and constant costs incurred by the normal form and lattice checking provide a straightforward extension of CKY-style context-free parsing for CCG. Komagata (1997) claims that the average complexity of CCG parsing is $O(n^3)$ even without the finite schematization of type raising (based on the parsing of 22 sentences consisting of around 20 words, with a lexicon of 200 entries and no derivation of semantics in the grammar; a morphological analyzer provided five analyses per second to the parser). Statistical techniques developed for lexicalized grammars (e.g., Collins 1997), readily apply to CCG to improve the average parsing performance in large-scale practical applications (Hockenmaier, Bierner, and Baldridge 2000). Both Collins and Hockenmeier, Bierner, and Baldridge used section 02-21 of the Wall Street Journal Corpus of Penn Treebank for training, which contains 40,886 words (70,151 lexical entries). A recent initiative (Oflazer, et al. 2001) aims to provide such a resource of around one million words for Turkish. It encodes in the Treebank surface-syntactic relations and the morphological breakdown of words. The latter is invaluable for training morphemic grammars and lexicons.

In morpheme-based parsing, lattice conditions help eliminate the permutation problem in endotypic categories. Such categories are typical of inflectional morphemes. For instance, assume that three morphemes m_1, m_2, and m_3 have endotypic categories (say $N\backslash N$), that they can appear only in this order, and that they are all optional. The categorization of m_i as $\overset{\kappa_i'}{\lhd} N\backslash \overset{\kappa_i}{\lhd} N$ such that $\kappa_i' \not\leq \kappa_i$ for all i, and $\kappa_{j-1}' \leq \kappa_j$ for $j = 1, 2, 3$ allows omissions (18a–b) but rules out the permutations (18c–d).[15]

(18) a. *stem* m_1 m_2 m_3

$\overset{\kappa_0}{\lhd} N$ $\quad \overset{\kappa_1'}{\lhd} N\backslash \overset{\kappa_1}{\lhd} N$ $\quad \overset{\kappa_2'}{\lhd} N\backslash \overset{\kappa_2}{\lhd} N$ $\quad \overset{\kappa_3'}{\lhd} N\backslash \overset{\kappa_3}{\lhd} N$

———————— <

$\overset{\kappa_1'}{\lhd} N$ because $\kappa_0 \leq \kappa_1$

———————— <

$\overset{\kappa_2'}{\lhd} N$ because $\kappa_1' \leq \kappa_2$

———————— <

$\overset{\kappa_3'}{\lhd} N$ because $\kappa_2' \leq \kappa_3$

b. *stem* m_3

———————— <

$\overset{\kappa_3'}{\lhd} N$ because $\kappa_0 \leq \kappa_3$

15 Three asterisks in the line indicate that the derivation is not licensed.

c. **stem* m_2 m_1 m_3

$\overline{\hspace{10em}}_{<}$

$\overset{\kappa'_2}{\triangleleft} N$ because $\kappa_0 \leq \kappa_2$

$\overline{\hspace{6em}{***}\hspace{6em}}_{<}$

$\kappa'_2 \not\leq \kappa_1$ because $\kappa_1 < \kappa'_1 \leq \kappa_2 < \kappa'_2$

d. **stem* m_1 m_3 m_2

$\overline{\hspace{10em}}_{<}$

$\overset{\kappa'_1}{\triangleleft} N$ because $\kappa_0 \leq \kappa_1$

$\overline{\hspace{12em}}_{<}$

$\overset{\kappa'_3}{\triangleleft} N$ because $\kappa'_1 \leq \kappa_3$

$\overline{\hspace{6em}{***}\hspace{6em}}_{<}$

$\kappa'_3 \not\leq \kappa_2$ because $\kappa_2 < \kappa'_2 \leq \kappa_3 < \kappa'_3$

The lattice and its consistency condition on derivability offer varying degrees of flexibility. A lattice with only $\top$ and the relation $\leq$ would undo all the effects of parameterization; it would be equivalent to a syntactic grammar in which every basic category X stands for $\overset{\top}{\triangleleft} X$. To enforce a completely lexemic syntax, a lattice with $\top$ and *free* would define all functional categories as functions over free forms.

Morphological processing seems inevitable for languages like Turkish, and morphological and lexical ambiguity such as that shown in (19) must be passed on to syntax irrespective of how inflectional morphology is processed (isolated from or integrated with syntax). For the verbal paradigm, Jurafsky and Martin (2000) reports Oflazer's estimation that inflectional suffixes alone create around 40,000 word forms per root. In the nominal paradigm, iterative processes such as *ki*-relativization (Section 6.5) can create millions of word forms per nominal root (Hankamer 1989).

(19) a. *kazma-ları*
pickaxe-POSS3p
'their pickaxe'

b. *kazma-lar-ı*
pickaxe-PLU-POSS3p
'their pickaxes'

c. *kazma-lar-ı*
pickaxe-PLU-POSS3s
'his/her pickaxes'

d. *kaz-ma-ları*
dig-SUB-AGR
'their digging'

The questions that need to be answered related to processing are *(1)* What should a (super)linear fragment of processing for morphology deliver to (morpho)syntax? and *(2)* Is the syntax lexemic or morphemic? The problems with lexemic syntax, which stem from mismatches with semantics, were highlighted in the introduction. In other

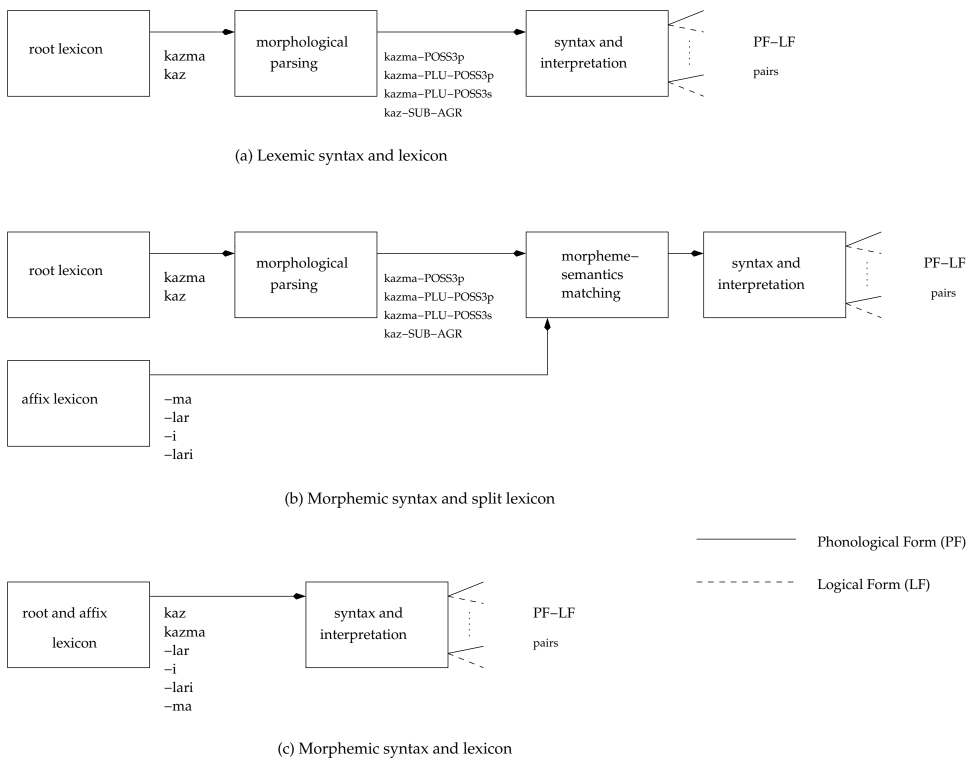


Figure 3
The processing of *kazmaları* in three different architectures (see Example (19) for glosses).

words, a lexemic grammar (e.g., Figure 3a) is computationally nontransparent when interpretation is a component of an NLP system.

Regarding the first question, let us consider two architectures from the perspective of the lexicon for the purpose of morphology, morphemic syntax, and semantics interface. The architecture in Figure 3b incorporates the current proposal as an interpretive front end to a morphological analyzer such as Oflazer's (1994), which delivers the analyses of words as a stream of morphemes out of which the bound morphemes have to be matched with their semantics from the affix lexicon to be interpretable in grammar. The advantage of this model is its efficiency; morphological parsing of words is—in principle—linear context free; hence, finite-state techniques and their computational advantages readily apply. But the uninterpretable surface forms of bound morphemes must match with those of the affix lexicon, and this is not necessarily a one-to-one mapping because of multiple lexical assignments for capturing syntactic–semantic distinctions (e.g., dative case as a direct object, indirect object, or adjunct marker or *-i* as a possessive and/or compound marker). Surface form–semantics pairing is not a trivial task, particularly in the case of lexically composite affixes, which require semantic composition as well as tokenization. The matching process needs to be aware of all the syntactic contexts in which certain affix sequences act as a unit, for example, relative participles and agreement markers (*-diğ-i* relative participle as -OP-POSS or -OP-AGR), possessive and compound markers, etc., for Turkish. The factorization of syntactic issues into a morphological analyzer would also make the separate morphological component nonmodular or expand its number of states to factor in these concerns (e.g., treating the -OP-POSS sequence as a state different from -OP followed

Table 2
Parsing performance.

Sample text type	Number of items in text			Average number of parses/grammatical input		Average CPU time per test (milliseconds)		
	tests	words	morphs	PAS check	Normal form parse	Unrestr.	PAS check	Normal form parse
Word order and case	58	216	384	1.26	3.68	39	39	30
Subordination	14	70	137	3.00	5.09	267	270	180
Relativization	23	130	232	2.04	2.32	796	783	266
Control verbs	33	147	291	1.42	3.34	166	163	137
Possessives and compounds	26	109	200	1.23	2.47	137	135	98
Adjuncts	14	57	100	1.12	4.87	89	88	72
-ki relatives	24	66	179	1.07	1.54	36	36	35

Note: CPU times are for a Sun UltraSparc-4 running SICStus Prolog; lexical items include stems and inflectional affixes.

by -POSS, in which -POSS is not interpreted with the semantics of possession but that of agreement marking). Not knowing how many of the syntactic distinctions are handled by the morphological analyzer, a subsequent interpreter may need to reconsult the grammar if scoping problems arise.

The architecture in Figure 3c describes the current implementation of the proposal. Bound morphemes are fed to the parser along with their interpretation. This model is preferred over that presented in Figure 3b for its simplicity in design and extendibility.[16] The price is lesser efficiency due to context-free processing of inflectional morphology. By one estimate (Oflazer, Gocmen, and Bozsahin 1994), Turkish has 59 inflectional morphemes out of a total of 166 bound morphemes, and Oflazer (personal communication) notes that the average number of bound morphemes per word in unrestricted corpora is around 2.8, including derivational affixes. In a news corpus of 850,000 words, the average number of inflections per word is less than two (Oflazer et al. 2001). This is tolerable for sentences of moderate length in terms of the extra burden it puts on the context-free parser. Table 2 shows the results of our tests with a Prolog implementation of the system on different kinds of constructions. The test cases included 10 lexical items on average, with an average parsing time of 0.32 seconds per sentence. A relatively long sentence (12 words, 21 morphemes) took 2.9 seconds to parse. The longest sentence (20 words, 37 morphemes) took 40 seconds. The lexicon for the experiment included 700 entries; 139 were free morphemes and 561 were bound morphemes compiled out of 105 allomorphic representations (including all the ambiguous interpretations of bound morphemes and the results of lexical rules). For a rough comparison with an existing NLP system with no disambiguation

16 The morphological analyzer would be in no better position to handle morpheme–semantics pairing if the architecture in Figure 3b were implemented with an integrated lexicon of roots and affixes. For instance, -POSS would still require distinct states because of the difference in the semantics of possession and agreement marking coming from the lexicon.

aids, Güngördü and Oflazer (1995) reported average parsing times of around 10 seconds per sentence for a lexicon of 24,000 free morphemes, and their morphological analyzer delivered around two analyses per second to a lexemic grammar. Oflazer's later (1996) morphological analyzer contained an abstract morphotactic component of around 50 states for inflections, which resulted in compilation to 30,000 states and 100,000 transitions when the morphophonemic rules were added to the system.

In conclusion, we note that the current proposal for a morphemic lexicon and grammar is compatible with both a separate morphological component (Figure 3b) and syntax-integrated inflectional morphology (Figure 3c). The architecture in Figure 3b may in fact be more suitable for inflecting languages (e.g., Russian) in which the surface forms of bound morphemes are difficult to isolate (e.g., *méste*, locative singular of *késto*) but can be delivered as a sequence of morpheme labels by a morphological analyzer (e.g. *mésto*-SING-LOC) to be matched with the lexical type assignments to -SING and -LOC for grammatical interpretation.

It might be argued that in computational models of the type in Figure 3b, the lattice is not necessary, because the morphological analyzer embodies the tactical component. But not only tactical problems (cf. Example (18) and its discussion) but also transparent scoping in syntax and semantics is regulated by the use of lattice in type assignments, and that is our main concern. We show examples of such cases in the remainder of the article. Thus the nonredundant role of the lattice decouples the morphemic grammar–lexicon from the kind of morphological analysis performed in the back end.

5. Case Study: The English Plural

In this section, we present a morphosyntactic treatment of the English plural morpheme. The lattice for English is shown in Figure 2b. We follow Carpenter (1997) in categorizing numerical modifiers and intersective adjectives as plural noun modifiers: *four boys* is interpreted as four(plu boy) and *green boxes* as green(plu box). This bracketing reflects the "set of sets" interpretation of the plural noun; four(plu boy) denotes the set of nonempty nonsingleton sets of boys with four members. The type assignments in (20) correctly interpret the interaction of the plural and these modifiers (cf. 21a–b). The endotypic category of the plural also allows phrase-internal number agreement for languages that require it; the agreement can be regulated over the category N before the specifier is applied to the noun group to obtain NP.

(20)

$$\begin{array}{lll} \text{-PLU} & := & \overset{a}{\circ} s - \overset{n}{\triangleleft} N \backslash \overset{b}{\triangleleft} N\colon \lambda x.\mathsf{plu}\,x \\ \text{four} & := & \overset{s}{\circ} \mathit{four} - \overset{n}{\triangleleft} N / \overset{n}{\bowtie} N\colon \lambda x.\mathsf{four}\,x \\ \text{green} & := & \overset{s}{\circ} \mathit{green} - \overset{n}{\triangleleft} N / \overset{n}{\triangleleft} N\colon \lambda x.\mathsf{green}\,x \end{array}$$

(21) a.

$$\begin{array}{ccc} \mathit{four} & \mathit{boy} & \mathit{\text{-}s} \\ \hline \overset{n}{\triangleleft} N / \overset{n}{\bowtie} N & \overset{b}{\triangleleft} N & \overset{n}{\triangleleft} N \backslash \overset{b}{\triangleleft} N \\ & \multicolumn{2}{c}{\underline{\hspace{8em}}_{<}} \\ & \multicolumn{2}{c}{\overset{n}{\triangleleft} N\colon\ \mathsf{plu\ boy}} \\ \multicolumn{3}{c}{\underline{\hspace{14em}}_{>}} \\ \multicolumn{3}{c}{\overset{n}{\triangleleft} N : \mathsf{four(plu\ boy)}} \end{array}$$

b. *four* *boy* *-s*

$$\underline{\quad\quad}^{***}\underline{\quad\quad}{>}\qquad \underline{\quad\quad}$$

$$\overset{n}{\triangleleft} N\text{: four boy} \qquad \overset{n}{\triangleleft} N\backslash \overset{b}{\triangleleft} N$$

because n-base $\neq$ n-num

$$\underline{\quad\quad\quad\quad}^{***}\underline{\quad\quad\quad\quad}$$

$$\overset{n}{\triangleleft} N\text{: * plu(four boy)}$$

Carpenter (1997) points out that nonintersective adjectives (e.g, *toy, fake, alleged*) are unlike numerical modifiers and intersective adjectives in that their semantics requires phrasal (wide) scope for -PLU, corresponding to the "set of things" interpretation of the plural noun. Thus, *toy guns* is interpreted as plu(toy gun) because the plural outscopes the modification. It denotes a nonempty nonsingleton set of things that are not really guns but toy guns. *toy(plu gun) would interpret plu over guns. The situation is precisely the opposite of (21); we need the second derivational pattern to go through and the first one to fail. The following category for nonintersective adjectives derives the wide scope for -PLU but not the narrow scope:

(22) toy := $\overset{s}{\circ}\, toy - \overset{b}{\triangleleft} N / \overset{b}{\triangleleft} N\text{: } \lambda x.\text{toy}\, x$

(23) a. *toy* *gun* *-s*

$$\underline{\quad\quad} \qquad \underline{\quad\quad\quad\quad}{<}$$

$$\overset{b}{\triangleleft} N / \overset{b}{\triangleleft} N \qquad \overset{n}{\triangleleft} N\text{: plu gun}$$

$$\underline{\quad\quad\quad\quad}^{***}\underline{\quad\quad\quad\quad}$$

$$\overset{n}{\triangleleft} N\text{ : *toy(plu gun)}$$

because n-num $\not\leq$ n-base

b. *toy* *gun* *-s*

$$\underline{\quad\quad} \qquad \underline{\quad} \qquad \underline{\quad\quad}$$

$$\overset{b}{\triangleleft} N / \overset{b}{\triangleleft} N \qquad \overset{b}{\triangleleft} N \qquad \overset{n}{\triangleleft} N\backslash \overset{b}{\triangleleft} N$$

$$\underline{\quad\quad\quad\quad}{>}$$

$$\overset{b}{\triangleleft} N\text{: toy gun}$$

$$\underline{\quad\quad\quad\quad\quad\quad}{<}$$

$$\overset{n}{\triangleleft} N\text{ : plu(toy gun)}$$

Carpenter (1997) avoided rebracketing because of the plural through lexical type assignments to plural nouns and a phonologically null lexical entry to obtain different semantic effects of the plural. In our formulation, there is no lexical entry for inflected forms and no phonologically null type assignment to account for the distinction in different types of plural modification; there is only one (phonologically realized) category for -PLU.[17] The modifiers differ only in the kind and degree of morphosyntactic control. Strict control ($\bowtie$) on *four* disallows *four boy*, and flexible control ($\triangleleft$) on *green* also handles *green box*. *Four green boxes* is interpreted as four(green(plu box)),

17 This is not to say that there is only one model-theoretic interpretation of plu. "Sets of sets" and "set of individuals" valuations of plu can be carried over the PAS.

not as *four(plu(green box)), and *four toy guns* is interpreted as four(plu(toy gun)), not as *plu(four(toy gun)). These derivations preserve the domain of the modifiers and the plural without rebracketing.

6. Case Study: Turkish Morphosyntax

There have been several computational studies to model morphology–syntax interaction in Turkish. These unification-based approaches represent varying degrees of integration. Güngördü and Oflazer (1995) isolates morphology from syntax by having separate modules (a finite-state transducer for the former, and an LFG component for the latter), that is, the syntax is lexemic. The morphological component is expected to handle all aspects of morphology, including inflections and derivations. In Sehitoglu and Bozsahin (1999), lexical rules implement inflectional morphology, and derivations are assumed to take place in the lexicon. Hoffman's (1995) categorial analysis of Turkish is also lexemic; all lexical entries are fully inflected. Interpretive components of these systems face the aforementioned difficulties because of their commitment to lexemic syntax. Inflectional morphology is incorporated into syntax in another categorial approach (Bozsahin and Göçmen 1995), but morphotactic constraints are modeled with nonmonotonic unification, such as nonexistence checks for features and overrides. The system cannot make finer distinctions in morphosyntactic types either. The result is an overgenerating and nontransparent integration of morphology and syntax because of the possibility of rebracketing and the unresolved representational basis of the lexicon.

In this section, we outline the application of the proposed framework to Turkish. We analyze a large fragment of the language, without any claims for a comprehensive grammar. The phenomena modeled here exhibit particular morphosyntactic problems described in the preceding sections. We assume the binding theory in Steedman (1996), which is predicated over the PAS. In each section, we provide a brief empirical observation about the phenomenon, propose lexical type assignments, exemplify derivations of the parser, and briefly discuss the constraints imposed by morphosyntactic types. Because of space considerations, we sometimes use abbreviated forms in derivations such as the genitive affix's $(N/(N\backslash N))\backslash N$ category for $(\overset{o}{\lhd} N/(\overset{o}{\bowtie} N_{pn}\backslash \overset{o}{\bowtie} N_{pn}))\backslash \overset{o}{\lhd} N_{pn}$, but the parser operates on full morphosyntactic representations.

6.1 Case Marking and Word Order

Turkish is regarded as a free constituent order language; all permutations of the predicate and its arguments are grammatical in main clauses, being subject to constraints on discourse and semantic properties such as definiteness and referentiality of the argument and topic–focus distinctions. The mapping of surface functions to grammatical relations is mediated by case marking. Word order variation has lesser functionality in embedded clauses because embedded arguments are less accessible to surface discourse functions like topic and focus. Embedded clauses are verb final.

6.1.1 Lexical Types. We start with the lexical type assignments for the verbs. We use the abbreviations in (24a) when no confusion arises about the arguments' case or morphosyntactic type. Verb-final orders are regarded as basic, which suggests the category $S\backslash NP\backslash NP$ for transitive verbs. But Janeway (1990) argued that such underspecification for verb-peripheral languages causes undesirable ambiguity. Grammatical relations of the arguments are determined not by directionality but by case in such languages. The category $S\backslash NP_{nom}\backslash NP_{acc}$ resolves the ambiguity (24b–c).

(24) a. $IV = \quad S\backslash NP$
$TV = \quad S\backslash NP\backslash NP$
$DV = \quad S\backslash NP\backslash NP\backslash NP$

b. sev (like) := $\overset{s}{\circ}\ sev - \overset{v}{\triangleleft} S\backslash \overset{f}{\triangleleft} NP_{nom}\backslash \overset{f}{\triangleleft} NP_{acc}\colon \lambda x.\lambda y.\mathsf{like}\, xy$

c. ver (give) := $\overset{s}{\circ}\ ver - \overset{v}{\triangleleft} S\backslash \overset{f}{\triangleleft} NP_{nom}\backslash \overset{f}{\triangleleft} NP_{dat}\backslash \overset{f}{\triangleleft} NP_{acc}\colon \lambda x.\lambda y.\lambda z.\mathsf{give}\, yxz$

d. $\overset{v}{\triangleleft} S\backslash \overset{f}{\triangleleft} NP_{nom}\backslash \overset{f}{\triangleleft} NP_{acc} : \lambda x.\lambda y.\mathsf{like}\, xy \Rightarrow$
$\overset{v}{\triangleleft} S\backslash \overset{f}{\triangleleft} NP^{+ref}_{acc}\backslash \overset{f}{\triangleleft} NP_{nom} : \lambda y.\lambda x.\mathsf{like}\, xy$

e. -ACC := $\overset{a}{\circ}\ i|ı|u|ü|yi|yı|yu|yü - \overset{c}{\triangleleft} N_{acc}\backslash \overset{o}{\triangleleft} N\colon \lambda f.f$

f. -LOC := $\overset{a}{\circ}\ de|da|te|ta - (\overset{\alpha}{\triangleleft} S/ \overset{\alpha}{\triangleleft} S)\backslash \overset{o}{\triangleleft} N\colon \lambda x.\lambda f.\mathsf{at}\, fx$

Gapping behavior seems to indicate that Turkish is verb final, not just SOV. SO and OS syntactic types must be distinguished to account for SO & SOV, OS & OSV, *SO & OSV and *OS & SOV. The OS & OSV pattern requires the lexical category $S\backslash NP_{acc}\backslash NP_{nom}$ for the verb (Bozsahin 2000b). SOV and OSV base orders can be captured uniquely in the lexicon in set-CCG notation as $S\backslash\{NP_{acc}, NP_{nom}\}$. Set-CCG is strongly equivalent to CCG (Baldridge 1999). We distinguish SOV and OSV lexically, however, because OSV requires referential objects (25a–b). OSV is generated from SOV by a lexical rule (24d). This is genuine lexical ambiguity, because the two related entries differ in semantics (referentiality).

(25) a. *Kitab-ı adam oku-du*
Book-ACC man.NOM read-TENSE
'The man read the book.'

b. **Kitap adam oku-du*
Book man.NOM read-TENSE

Regarding the relationship between case and the specifiers, it is questionable whether Turkish has a discernible syntactic category for determiners. There is no lexical functor that takes an *N* and yields an *NP*. The only article, the indefinite *bir ('*a*')*, makes a distinction in discourse properties (26). Specifying case as a determiner (e.g., $NP\backslash N$) does not alleviate the problem, either. Ignoring the problem of case stacking for a moment, zero marking of the surface subject and the indefinite object takes us back to where we started.

(26) *Çocuk yeşil bir elma/elma/elma-yı ye-miş*
child.NOM green an apple/apple/apple-ACC eat-TENSE
'The child ate a green apple.' (indefinite but referential apple)
'The child ate green apple.' (indefinite and nonreferential apple)
'The child ate the green apple.' (definite and referential apple)

Making the nouns lexically ambiguous (N or NP) would also require that all functions onto nouns be ambiguous ($N\backslash N$ and $NP\backslash NP$ for inflections, N/N and NP/NP for adjectives, etc.). Redundancy of this kind in the lexicon is not desirable, since it is introduced purely for formal reasons with no distinction in meaning. We accommodate these concerns by positing a special case of type raising for Turkish (27). Similarly, contraposition turns Ns into functors looking for NPs.

(27) Type Raising for Turkish:
$$N_{agr} : a \;\Rightarrow\; T/(T\backslash \overset{f}{\triangleleft} NP_{agr}) : \lambda f.f[a]$$
$$\;\Rightarrow\; T\backslash(T/ \overset{f}{\triangleleft} NP_{agr}) : \lambda f.f[a]$$
$$T \in \{S, S\backslash NP, S\backslash NP\backslash NP, S\backslash NP\backslash NP\backslash NP\}$$

The noun that is type raised can be a syntactically derived noun (28). SO (and OS) constituency required for gapping is provided by >T and >B.

(28)

Mehmet	*küçük yeşil kitab-ı,*	*çocuk da*	*yeni gelen dergi-yi*	*oku-du*
M.NOM	little green book-ACC	child-COORD	new come mag.-ACC	read-TENSE
N_{nom}	N_{acc}	N_{nom}	N_{acc}	$S\backslash NP_{nom}\backslash NP_{acc}$
>T: $S/(S\backslash NP_{nom})$	>T: $(S\backslash NP)/(S\backslash NP\backslash NP_{acc})$	⋮		

>B (Mehmet … kitab-ı): $S/(S\backslash NP_{nom}\backslash NP_{acc})$

>B (çocuk da … dergi-yi): $S/(S\backslash NP_{nom}\backslash NP_{acc})$

& : $S/(S\backslash NP_{nom}\backslash NP_{acc})$

> : S

'Mehmet read the little green book, and the child, the newly arrived magazine'.

Our lexical type assignment to case morphemes (24e–f) departs from other CCG analyses of case (e.g., Steedman 1985, 1991a, Bozsahin 1998). These studies correlate morphological case with type raising of arguments, in the case of Bozsahin (1998), via a value-raised category assignment to case morphemes. Evidence from NP-internal case agreement and case stacking (Kracht 1999) challenges the type-raising approach. Agreement phenomena require that case (which can be marked on articles, adjectives, and nouns) be regulated as an agreement feature within the category N *before* the case-marked argument looks for the verb via type raising. Kracht observes that, in case stacking, there may be other morphemes between two case morphemes. Thus, treating the two cases as composite affixes for the purpose of type raising is not feasible. If the first case type-raises the noun to say, $T/(T\backslash NP)$, the second case would require a category, $(T/(T\backslash NP))\backslash(T/(T\backslash NP))$; that is, it is endotypic. Hence, an endotypic category for case (like other inflections in the paradigm) subsumes the type-raising analysis of case provided that type raising is available in the grammar, not necessarily anchored to case.

We analyze case as an endotypic functor of type $N\backslash N$ (24e)—hence allow for phrase-internal agreement for languages that require it and provide type raising in grammar as in (27). Abandoning the type-raising analysis of case does not necessitate taking liberties in the directionality of the categories, such as the use of nondirectional slash (|) in multiset-CCG (Hoffman 1995). Contraposition and type raising in grammar can account for free word order and gapping facts with fully directional syntactic types (Bozsahin 2000a).

6.1.2 Derivations. The wide scope of case is captured by treating its argument type as non-case-marked N ($\overset{o}{\triangleleft} N$) and the type of noun modifiers as functions onto non-case-marked nouns of a particular domain, for example, $\overset{b}{\triangleleft} N$ for nonintersective adjectives and $\overset{n}{\triangleleft} N$ for intersective adjectives (29a). The same strategy in type assignments to other nominal inflections allows them to outscope nominal modification, for example, (29b).

(29) a.

Mehmet	[[*oyuncak*	*araba*]	*-lar*]	*-ı*	*sev-er*
M.NOM	toy	car	-PLU	-ACC	like-TENSE
$\overset{b}{\triangleleft} N_{nom}$	$\overset{b}{\triangleleft} N / \overset{b}{\triangleleft} N$	$\overset{b}{\triangleleft} N$	$\overset{n}{\triangleleft} N \backslash \overset{b}{\triangleleft} N$	$\overset{c}{\triangleleft} N_{acc} \backslash \overset{o}{\triangleleft} N$	$\overset{t}{\triangleleft} S \backslash \overset{f}{\triangleleft} NP_{nom} \backslash \overset{f}{\triangleleft} NP_{acc}$ (<B)
: mehmet	: $\lambda x.$toy x	: car	: $\lambda x.$plu x	: $\lambda f.f$	: $\lambda x.\lambda y.$like xy
>T $S/(S\backslash \overset{f}{\triangleleft} NP_{nom})$: $\lambda f.f$[mehmet]					

$$\overset{b}{\triangleleft} N\text{: toy car} \quad (>)$$

$$\overset{n}{\triangleleft} N\text{: plu(toy car)} \quad (<)$$

$$\overset{c}{\triangleleft} N_{acc}\text{: plu(toy car)} \quad (<)$$

$$(S\backslash NP)/(S\backslash NP\backslash \overset{f}{\triangleleft} NP_{acc})\text{: } \lambda g.g[\text{plu(toy car)}] \quad (>T)$$

$$\overset{t}{\triangleleft} S\backslash \overset{f}{\triangleleft} NP_{nom}\text{: } \lambda y.\text{like(plu(toy car))}y \quad (>)$$

$$S\text{: like(plu(toy car))mehmet} \quad (>)$$

'Mehmet likes toy cars.'

b. *Adam-ın* [*küçük kırmızı araba*]-*sı*
Man-GEN little red car-POSS
'the man's little red car' = poss(little(red car))man

A word-based alternative for reconciling the semantic (wide) scope of inflections and their morphological (narrow) attachment to stems runs into difficulties even if we assume that morphemes carry type assignments—and hence have representational status—but that they always combine with stems first. We use syntactic types to show the problem. If -PLU and -ACC in (29a) combine with the stem first, only the narrow-scope reading of the plural and case is possible (30a). Plu(toy car) is not derivable with word-based modification. The morphosyntactic categories, however, are transparent to the scope of nominal modification (cf. (29a) and (30b)).

(30) a. *oyuncak* [[*araba*] *-lar*] *-ı*

toy car -PLU -ACC

N/N : $\lambda x.$toy x N : car $N\backslash N$: $\lambda x.$plu x $N_{acc}\backslash N$: $\lambda f.f$

$<$ N: plu car

$<$ N_{acc}: plu car

$>$ N: $*$ toy(plu car)

b. [*yeşil* [*araba*] *-lar*] *-ı*

green car -PLU -ACC

$\overset{n}{\triangleleft} N/\overset{n}{\triangleleft} N$: $\lambda x.$green x $\overset{b}{\triangleleft} N$: car $\overset{n}{\triangleleft} N\backslash\overset{b}{\triangleleft} N$: $\lambda x.$plu x $\overset{c}{\triangleleft} N_{acc}\backslash\overset{o}{\triangleleft} N$: $\lambda f.f$

$<$ $\overset{n}{\triangleleft} N$: plu car

$>$ $\overset{n}{\triangleleft} N$: green(plu car)

$<$ $\overset{c}{\triangleleft} N_{acc}$: green(plu car)

Surface case annotations on categories enable the grammar to capture the correct PAS in all permutations of S, O, and V while maintaining the discourse-relevant distinctions (31). Verb-final subordinate clauses are enforced by the directionality of the subordination morphemes in the lexicon.

(31) a. S O V

$>_T$ $S/(S\backslash\overset{f}{\triangleleft} NP_{nom})$ $>_T$ $(S\backslash NP)/(S\backslash NP\backslash\overset{f}{\triangleleft} NP_{acc})$ $S\backslash NP_{nom}\backslash NP_{acc}$

$>$ $S\backslash NP_{nom}$

$>$ S

b. O S V

$>_T$ $S/(S\backslash\overset{f}{\triangleleft} NP_{acc})$ $>_T$ $(S\backslash NP)/(S\backslash NP\backslash\overset{f}{\triangleleft} NP_{nom})$ $S\backslash NP_{acc}\backslash NP_{nom}$

$>$ $S\backslash NP_{acc}$

$>$ S

c. O V S

O: $(S\backslash NP)/(S\backslash NP\backslash \overset{f}{\triangleleft} NP_{acc})$ (>T) V: $S\backslash NP_{nom}\backslash NP_{acc}$ S: $S_{-t}\backslash(S\backslash NP_{nom})$ (>XP)

O V (>): $S\backslash NP_{nom}$

O V S (<): S_{-t}

d. S V O

S: $(S\backslash NP)/(S\backslash NP\backslash \overset{f}{\triangleleft} NP_{nom})$ (>T) V: $S\backslash NP_{acc}\backslash NP_{nom}$ O: $S_{-t}\backslash(S\backslash NP_{acc})$ (>XP)

S V (>): $S\backslash NP_{acc}$

S V O (<): S_{-t}

e. V S O

V: $S\backslash NP_{nom}\backslash NP_{acc}$ S: $S_{-t}\backslash(S\backslash NP_{nom})$ (>XP) O: $S_{-t}\backslash(S_{-t}\backslash NP_{acc})$ (>XP)

V S (<B): $S_{-t}\backslash NP_{acc}$

V S O (<): S_{-t}

f. V O S

V: $S\backslash NP_{acc}\backslash NP_{nom}$ O: $S_{-t}\backslash(S\backslash NP_{acc})$ (>XP) S: $S_{-t}\backslash(S_{-t}\backslash NP_{nom})$ (>XP)

V O (<B): $S_{-t}\backslash NP_{nom}$

V O S (<): S_{-t}

6.2 Subordination

Subordinate clauses can be classified as unmarked clauses (32a), infinitival clauses (32b), verbal nouns (32c), and nominalizations (32d). The latter two types require a genitive embedded subject, which agrees with the subordinate verb.

(32) a. *Mehmet* [*çocuk* *ev-e* *git-ti*] *san-dı*
M.NOM child.NOM house-DAT go-TENSE assume-TENSE
'Mehmet assumed that the child went home.'

b. *Çocuk* [*kız-a* *kalem-i* *ver-me*] *-yi* *unut-tu*
child.NOM girl-DAT pen-ACC give-SUB1i -ACC forget-TENSE
'The child forgot to give the pen to the girl.'

c. [*Çocuğ-un* *araba-da* *uyu-ma-sı*] *Mehmet'i* *kız-dır-dı*
child-GEN car-LOC sleep-SUB1g-POSS M-ACC anger-CAUS-TENSE
'Child's sleeping in the car made Mehmet angry.'

d. *Deniz* [*çocuğ-un* *uyu-duğ-u*] *-na* *inan-m-ıyor*
D.NOM child-GEN sleep-SUB2g-POSS -DAT believe-NEG-TENSE
'Deniz does not believe the child's sleeping.'

(33) a. *Deniz*$_i$ [*kendisi-nin*$_i$ *uyu-ma-dığ-ı*]*-nı* *söyle-di*
D.NOM self-GEN sleep-NEG-SUB2g-POSS-ACC2 say-TENSE
'Deniz$_i$ said that he$_i$ did not sleep.'

b. **kendisi*$_i$ [*Deniz'in*$_i$ *uyu-ma-dığ-ı*]*-nı söyle-di*

c. *Deniz*$_i$ *adam-ı*$_j$ [*kendi*$_{i/j}$ *arkadaş-ı-nın gör-düğ-ü*]*-ne* *inan-ıyor*
D.NOM man-ACC self friend-POSS see-SUB2g-POSS-DAT2 believe-TENSE
'Deniz$_i$ believes that his$_{i/j}$ friend saw the man$_j$.'

d. *Deniz*$_i$ *adam-a*$_j$ [*kendi*$_{i/*j}$ *kitab-ı-nı* *oku-duğ-u*]*-nu* *söyle-di*
D.NOM man-DAT self book-POSS-ACC2 read-SUB2g-POSS-ACC2 say-TENSE
'Deniz$_i$ told the man$_j$ that he read his$_{i/*j}$ book.'

6.2.1 Lexical Types. The asymmetries in (33) show that the obliqueness order in binding relations is preserved in subordination. This suggests the following bracketing, in which the embedded clause's position in the PAS of the matrix predicate is determined by its grammatical function.

Matrix-Pred ... *Matrix-Argument* ... *Embedded-Clause* ... *Matrix-Argument*

(34) -SUB1i (-ma) (infinitive) := $\overset{a}{\circ}\ ma - \overset{b}{\triangleleft} N\backslash(\overset{a}{\triangleleft} S\backslash \overset{f}{\triangleleft} NP_{nom})\colon \lambda f.f$

-SUB1g (-masɪ) (verbal noun) := $\overset{a}{\circ}\ mas\imath - \overset{o}{\triangleleft} N\backslash \overset{f}{\triangleleft} NP_{agr}\backslash(\overset{a}{\triangleleft} S\backslash \overset{f}{\triangleleft} NP_{nom})\colon \lambda f.f$

-SUB2g (-dığı) (nominalization) := $\overset{a}{\circ}\ d\imath\breve{g}\imath - \overset{o}{\triangleleft} N_{\mathrm{case=obl}}\backslash \overset{f}{\triangleleft} NP_{agr}\backslash(\overset{a}{\triangleleft} S\backslash \overset{f}{\triangleleft} NP_{nom})\colon \lambda f.f$

The wide scope of case markers on subordinate clauses implies that the subordinate markers themselves must have phrasal scope as well. Since case is a nominal inflection, the category of a subordinate marker must be a function onto *N*. Its argument is *IV* for infinitives and $NP_{agr}\backslash IV$ for others, which require genitive subjects (34). This yields two families of functors for subordination. The verb-final characteristics of the embedded clauses is ensured by the backward-looking main functor of the subordinate marker.

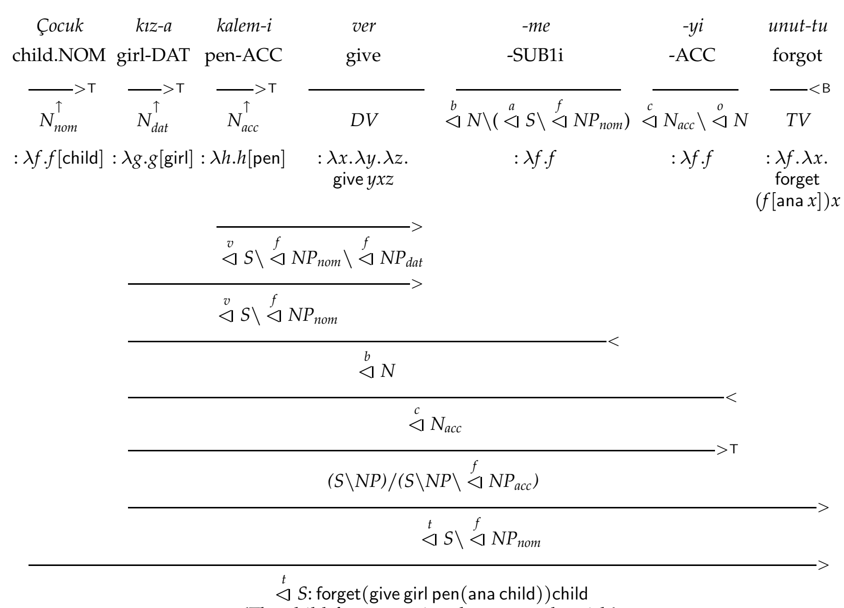
For morphosyntactic modality, the resulting nominalized predicate can receive only case, hence it has $\overset{o}{\triangleleft} N$ control. Verbal nouns refer to actions, and nominalizations refer to facts. Subordinate markers for the former are tenseless. A subordinate marker replaces the tense of the subordinate verb in nominalizations, yielding $\overset{a}{\triangleleft} S$

control on the verb. For subject raising, the result may undergo any nominal inflection ($\overset{b}{\triangleleft} N$).

Word order variation within the subordinate clause is constrained by the subject on the left and the verb on the right. This constraint is achieved by categorizing the embedded subjects as NP_{agr} and having a result category of N for all subordinate markers. If there were any contraposed element NP in the embedded clause, the category of the clause would be $S\backslash NP$, and the clause could not combine with the contraposed category such as $S_{-t}\backslash(S\backslash NP)$ on the right because the extraction category combines with a subordinate marker first, which is onto N, not $S\backslash NP$, hence composition (<B) could not take place.

6.2.2 Derivations. Example (35a) is the derivation of subject raising (we use $N^{\uparrow}$ as an abbreviation for a type-raised N when space is limited). We use Steedman's (1996) ana function to denote the binding of the embedded subject. Infinitive -SUB1i has phrasal scope in this example; the DV must be reduced to an IV before the infinitive can apply. Hence the subordination of intransitive clauses is only a special case in which the morphological scope of the infinitive works without rebracketing. Subject raising and coindexation with the matrix subject are made explicit in the raising category of *unut*. The systematic relationship between the raised and nonraised category of such verbs can be captured by a lexical rule, for example, $TV\colon \lambda x.\lambda y.\mathsf{forget}\, xy \Rightarrow TV\colon \lambda f.\lambda y.\mathsf{forget}(f[\mathsf{ana}\, y])y$.

(35b–c) contrast subject and nonsubject nominalizations. The difference is captured with the case distinction of the result type ($\overset{o}{\triangleleft} N$) for -SUB1g and -SUB2g. These examples also show the possibility of affix composition in the lexicon. For instance, we write *-ması* in (35b), which marks subordination and agreement together, instead of *-ma-sı*. Otherwise, *-ma* (SUB1g) would have to look to the right as a functor to enforce agreement, and the verb-final property of subordination could not be assured.

(35) a.


'The child forgot to give the pen to the girl.'

b. *Çocuğ-un uyu -ması Mehmet'i kızdır-dı*
child-GEN sleep -SUB1g M-ACC anger-TENSE

$\overline{NP_{agr}}^{>T}$ $\overline{IV}$ $\overline{N\backslash NP_{agr}\backslash IV}$ $\overline{IV/TV}^{>T}$ $\overline{TV}^{<B}$

$\overline{N\backslash NP_{agr}}^{<}$ $\overline{IV}^{>}$

$\overline{N}^{<}$

$\overline{S/IV}^{>T}$

$\overline{S\text{: anger(sleep child)mehmet}}^{>}$

'The child's sleeping angered Mehmet.'

c. **Çocuğ-un uyu-duğu Mehmet'i kızdır-dı*
sleep-SUB2g

6.3 The Morphosyntax of Control

The control verb's controlled argument is marked by the infinitive *-ma*, and the resulting nominalized embedded clause can undergo nominal inflections (36a–b). The infinitive *-ma* has the lexical type in (34). A potential conflict between an object control verb's subcategorization and PAS is resolved by case decoration: *zorla* 'force' and *tavsiye et* 'recommend' differ in their case requirements and what is controlled (36b–c). *tavsiye et*'s infinitive complement is accusative, whereas *zorla*'s is dative.

(36) a. *Çocuk [kitab-ı oku-ma]-ya çalış-tı*
child.NOM book-ACC read-SUB1i-DAT try-TENSE
'The child$_i$ tried [to $\_\__i$ read the book].'

b. *Mehmet çocuğ-u [kitab-ı oku-ma]-ya zorla-dı*
M.NOM child-ACC book-ACC read-SUB1i-DAT force-TENSE
'Mehmet$_j$ forced the child$_i$ [to $\_\__{i/*j}$ read the book].'

c. *Mehmet çocuğ-a/*-u [kitab-ı oku-ma]-yı/*-ya tavsiye et-ti*
M.NOM child-DAT/ACC book-ACC read-SUB1i-ACC/DAT recommend-TENSE
'Mehmet recommended the child$_i$ [to $\_\__i$ read the book].'

6.3.1 Lexical Types. Subject control verbs (e.g., *çalış* 'try'; *söz ver* 'promise') and object control verbs (e.g., *zorla*; *tavsiye et*) have the control property indicated in their PAS (37). The nonraising variety of these verbs is obtained via a lexical rule.

(37)
çalış := $\overset{s}{\circ}$ *çalış* $-$ $TV\text{: }\lambda q.\lambda z.\mathsf{try}(q[\mathsf{ana}\ z])z$
söz ver := $\overset{s}{\circ}$ *söz ver* $-$ $DV\text{: }\lambda q.\lambda z.\lambda w.\mathsf{promise}\ z(q[\mathsf{ana}\ w])w$
zorla := $\overset{s}{\circ}$ *zorla* $-$ $DV\text{: }\lambda z.\lambda q.\lambda w.\mathsf{force}(q[\mathsf{ana}\ z])zw$
tavsiye et := $\overset{s}{\circ}$ *tavsiye et* $-$ $DV\text{: }\lambda z\lambda q\lambda w.\mathsf{recommend}(q[\mathsf{ana}\ z])zw$

6.3.2 Derivations. The types in (37), coupled with the raising category of the infinitive, yield the derivations in (38). These examples compose the infinitive complement before a case can be applied on the nominalized predicate. This is possible because of the phrasal scope of *-ma* and the case markers. (38b) shows that although there may be two accusative-marked NPs, the arguments of the infinitive complement are identifiable; the *IV* scope of *-ma* implies that any (di)transitive subordinate verb must find its nonsubject arguments before the matrix verb gets its arguments. This type assignment strategy handles word order variations inside the infinitive complement and the matrix clause transparently.

(38) a. *Çocuk* *kitab-ı* *oku* *-ma* *-ya* *çalış-tı*
child.NOM book-ACC read -SUB -DAT try-TENSE

S/IV (>T) IV/TV (>T) TV $N\backslash IV$ $N_{dat}\backslash N$ TV (<B)
IV (>)
N (<)
N_{dat} (<)
IV/TV (>T)
IV (>)
S: try(read book(ana child))child (>)
'The child tried to read the book.'

b. *Mehmet* *çocuğ-u* *kitab-ı* *oku* *-ma* *-ya* *zorla-dı*
M.NOM child-ACC book-ACC read -SUB -DAT force-TENSE

S/IV (>T) IV/TV (>T) IV/TV (>T) TV $N\backslash IV$ $N_{dat}\backslash N$ DV (<B)
IV (>)
N (<)
N_{dat} (<)
TV/DV (>T)
TV (>)
IV (>)
S: force(read book(ana child))child mehmet (>)
'Mehmet forced the child to read the book.'

6.4 Relativization

There are two strategies for forming relative clauses: the subject participle strategy (SP) and the nonsubject participle strategy (OP). SP is realized by the affixes *-(y)An*, *-(y)AcAk*, and *-mIş*, and OP by *-dIk-* and *-(y)AcAk-*. OP triggers agreement similar to that of possessive constructions between the subject and the predicate of the relative clause (39b).

(39) a. *kitab-ı oku-yan adam*
book-ACC read-SP man
'the man that read/reads the book'

b. *adam-ın oku-duğ-u kitap*
man-GEN(AGR) read-OP-POSS(AGR) book
'the book that the man read'

6.4.1 Lexical Types. The categories in (40) make explicit the unbounded nature of relativization; type raising and composition can combine an indefinitely large sequence of constituents onto $S\backslash NP$.

(40) -SP := $\overset{a}{\circ}\, yan - (N^{\uparrow}/\overset{f}{\triangleleft} N)\backslash(\overset{a}{\triangleleft} S\backslash\overset{f}{\triangleleft} NP_{nom})$
$: \lambda P.\lambda x.\lambda Q.\mathsf{and}(Q[x])(P[x])$

-OP.AGR (argument) := $\overset{a}{\circ}\, d\imath\breve{g}\imath - (N^{\uparrow}/\overset{f}{\triangleleft} N)\backslash(\overset{a}{\triangleleft} S\backslash\overset{f}{\triangleleft} NP_{case=obl})$
$: \lambda P.\lambda x.\lambda Q.\mathsf{and}(Q[x])(P[x])$

-OP.AGR (adjunct) := $\overset{a}{\circ}\, d\imath\breve{g}\imath - (N^{\uparrow}/\overset{f}{\triangleleft} N)\backslash\overset{a}{\triangleleft} S$
$: \lambda P.\lambda x.\lambda Q.\mathsf{and}(Q[x])(\mathsf{at}(P[x])x)$

We present a formulation of relativization without any use of empty categories, traces, or movement. We follow the Montagovian treatment of relative clauses as noun restrictors of the semantic type $\lambda P.\lambda Q.\mathsf{and}(Q[x])(P[x])$, where P is the semantics of the relative clause and Q is the semantics of the predicate taking the relativized noun (x) as the argument. Montagovian analysis assumes a generalized quantifier (GQ) category for the determiner; that is, NP is the functor and VP is the argument. The determiner takes the relativized noun (and its semantically type-raised category) as an argument as well. In a language with determiners, the functor category of the overall NP can be made explicit by lexically value-raising the determiner with GQ semantics from, for example, NP/N to $(S/(S\backslash NP))/N = (S/VP)/N$. To achieve the same effect in a language that lacks determiners, we make NP the functor by lexically value-raising the relative participle from $(N/N)\backslash(S\backslash NP)$ to $(N^{\uparrow}/N)\backslash(S\backslash NP)$, in which $N^{\uparrow}/N$ denotes a value-raised noun, since $N^{\uparrow}$ is a type-raised category. The category of the relative participle unfolds to $((S/(S\backslash NP))/N)\backslash(S\backslash NP)$ and $(((S\backslash NP)/(S\backslash NP\backslash NP))/N)\backslash(S\backslash NP)$.

Relativization is strictly head final in Turkish. This implies that all relative participles are backward-looking functors that differ only in case requirements (cf. English relatives, which require different directionality, e.g., $(N\backslash N)/(S\backslash NP)$ for subjects and $(N\backslash N)/(S/NP)$ for nonsubjects). For morphosyntactic modality, the head noun has flexible control ($\overset{f}{\triangleleft}$ *N*), because any further grammatical marking on the head must be shared (41).

(41) *Adam-ın gör-düğ-ü çocuk-lar uyu-du*
man-GEN see-OP-POSS child-PLU sleep-TENSE
'The children that the man saw slept.' = and(sleep(plu child))(see(plu child)man)
≠*and(sleep(plu child))(see child man)

Morphologically, the agreement marker -POSS in OP strategy is a function over the -OP morpheme, but syntactically, the -OP morpheme triggers the agreement in the relative clause. Hence -OP-POSS can be treated as a lexically composite affix and glossed as -OP.AGR. This also ensures the verb-final property of the relativized clause by not positing a rightward-looking functor for -OP. As for attachment modality, relative participles are bound morphemes that are affixed to the predicate.

6.4.2 Derivations. (42a–d) show example derivations for subject, object, indirect object, and adjunct relativization. All nonsubject arguments are handled by a single -OP type (42b–c). Relativizing the specifier of an argument uses the same strategy as the argument. This phenomenon calls for another well-regulated lexical assignment schema, for example, $(N^{\uparrow}/N)\backslash(N\backslash N)\backslash IV$ for the relativized specifier of the subject. (42e) is an example of relativizing the subject's specifier. Configurationality within the noun group is maintained by backward directionality of the categories.

(42) a. *kitab-ı oku -yan adam uyu-du*
book-ACC read -SP man sleep-TENSE

$IV/TV : \lambda f.f[book]$ ($>_T$) $\quad TV : \lambda x.\lambda y.\mathsf{read}\,xy \quad (N^{\uparrow}/N)\backslash IV : \lambda P.\lambda x.\lambda Q.\mathsf{and}(Q[x])(P[x]) \quad N : \mathsf{man} \quad IV : \lambda x.\mathsf{sleep}\,x$ ($<_B$)

$>$ $IV: \lambda y.\mathsf{read\ book}\,y$

$<$ $N^{\uparrow}/N: \lambda x.\lambda Q.\mathsf{and}(Q[x])(\mathsf{read\ book}\,x)$

$>$ $N^{\uparrow}{=}S/(S\backslash NP): \lambda Q.\mathsf{and}(Q[\mathsf{man}])(\mathsf{read\ book\ man})$

$>$ $S: \mathsf{and}(\mathsf{sleep\ man})(\mathsf{read\ book\ man})$

'The man who read the book slept.'

b. *adam-ın gör -düğü çocuk uyu-du*
man-GEN read -OP.AGR child sleep-TENSE

NP_{agr} ($<$) $\quad TV \quad (N^{\uparrow}/N)\backslash IV_{agr} \quad N \quad IV$ ($<_B$)

$<$ IV_{agr}

$<$ $N^{\uparrow}/N$

$<$ $N^{\uparrow}{=}S/IV$

$>$ $S: \mathsf{and}(\mathsf{sleep\ child})(\mathsf{see\ child\ man})$

'The child whom the man saw slept.'

c. *çocuğ-un* *kitab-ı* *ver* *-diği* *adam* *uyu-du*
child-GEN book-ACC give -OP.AGR man sleep-TENSE

NP_{agr} (<) TV/DV (>T) DV $(N^{\uparrow}/N)\backslash IV_{agr}$ N IV (<B)

TV (>)

IV_{agr} (<)

$N^{\uparrow}/N$ (<)

$N^{\uparrow}{=}S\backslash IV$ (>)

S: and(sleep man)(give man book child) (>)

'The man to whom the child gave the book slept.'

d. *çocuğ-un* *uyu* *-duğu* *araba* *bozul-du*
child-GEN sleep -OP.AGR car break-TENSE

NP_{agr} (<) IV $(N^{\uparrow}/N)\backslash S$ N IV (<B)

S (<)

$N^{\uparrow}/N$ (<)

$N^{\uparrow}{=}S/IV$ (>)

S: and(break car)(at(sleep child)car) (>)

'The car that the child slept in broke.'

e. *çocuğ* *-u* *uyu* *-yan* *adam* *kız-dı*
child -POSS sleep -SP man anger-TENSE

N $N\backslash N\backslash N$ IV $(N^{\uparrow}/N)\backslash(N\backslash N)\backslash IV$ N IV (<B)

$N\backslash N$ (<) $(N^{\uparrow}/N)\backslash(N\backslash N)$ (<)

$N^{\uparrow}/N$ (<)

$N^{\uparrow}{=}S/IV$ (>)

S: and(sleep(poss child man))(anger man) (>)

'The man whose child slept got angry.'

As these examples indicate, -SP and -OP do not range over the verb stem in semantic scope; they cover the entire relative clause. The wide scope of -SP and -OP resolves the inconsistency pointed out in the introduction (5b–c), which was mainly due to coindexation in unification accounts and the lexemic nature of the lexicon. Isolating the relative participle inflections in a morphological component undermines the transparency of derivations. Note also that -OP is categorially transparent to the arity of the verb; a *DV* must be reduced to an *IV* before -OP applies to the verb complex (42c). This is possible only when -OP has phrasal scope.

6.5 *Ki*-relativization

Ki-relativization is a morphosyntactic process that can generate indefinitely long words of relative pronouns and relative adjectives. *-ki* can be attached to case-marked nouns whose case relation is one of possession, time, or place (i.e., the genitive and the locative). Its effect is to create a nominal stem on which all inflections can start again (43a–b). It produces relative pronouns (43c) and relative adjectives (43d) with the locative and relative pronouns with the genitive.

(43) a. *araba-da-ki*
car-LOC-REL
'the one in the car'

b. *çocuğ-un ev-i-nde-ki-ler-in-ki*
child-GEN house-POSS-LOC2-REL-PLU-GEN-REL
lit. 'The one that belongs to the ones that are in the child's house'

c. *Ben ev-de-ki-ni hiç kullan-ma-dı-m*
I.NOM house-LOC-REL-ACC2 never use-NEG-TENSE-PERS.1s
'I never used the one at home.'

d. *ev-de-ki hediye*
house-LOC-REL present
'the present$_i$, the one$_i$ at home'

6.5.1 Lexical Types.

(44) a. -PROki (locative) := $\overset{a}{\circ}\ ki - \overset{l}{\bowtie} N \backslash \overset{c}{\bowtie} N_{loc}\colon \lambda x.\lambda f.\mathsf{and}(\mathsf{at\ PRO}\ x)(f[\mathsf{PRO}])$

b. -ADJki := $\overset{a}{\circ}\ ki - (\overset{l}{\bowtie} N / \overset{n}{\triangleleft} N) \backslash \overset{c}{\bowtie} N_{loc}\colon \lambda x.\lambda y.\lambda f.\mathsf{and}(\mathsf{at}\ xy)(f[y])$

c. -PROki (genitive) := $\overset{a}{\circ}\ ki - \overset{l}{\bowtie} N \backslash N^{\uparrow}_{gen}\colon \lambda x.\lambda f.\mathsf{and}(\mathsf{poss\ PRO}\ x)(f[\mathsf{PRO}])$

d. sabahki := $\overset{s}{\circ}\ sabahki - \overset{l}{\bowtie} N / \overset{n}{\triangleleft} N\colon \lambda x.\lambda f.\mathsf{and}(\mathsf{at\ morning}\ x)(f[x])$

e. ki (that) := $\overset{c}{\circ}\ ki - (N^{\uparrow} \backslash \overset{f}{\triangleleft} N)/(\overset{v}{\triangleleft} S \backslash \overset{f}{\triangleleft} NP_{nom})$
$: \lambda P.\lambda x.\lambda Q.\mathsf{and}(Q[x])(P[x])$

$N^{\uparrow}_{gen}$ is a shorthand for the *N/(N\N)* category of a type-raised genitive. In (43c), pronominal *one* (PRO) cannot be bound to *ev* (44a). Adjectival interpretation (43d) associates the relative adjective with the relativized noun (44b). For morphosyntactic modality, *ki*-marked nouns behave like possessive-marked nouns in case marking, which requires strict control over the possessive ($\overset{o}{\bowtie} N$). This presents a dilemma: Morphologically, *-ki* creates a nominal stem that can undergo all nominal inflections again, but, as (45a) indicates, the stem does not take the CASE (ACC, DAT, etc.) that is common to nouns unmarked on the possessive. Thus CASE2 in (45c) must refer to another diacritic (n-relbase, or $\overset{l}{\bowtie}$) to eliminate (45b). This diacritic controls the result category of *-ki*. The value-raised varieties of (44a–c) are assigned a type similar to the type of relative participles. Inherently temporal nouns such as *sabah* ('morning') can take *-ki* without the locative. They can be lexicalized without overgeneration with the help of the morphosyntactic modality $\overset{l}{\bowtie}$ (44d).

(45) a. **ev-de-ki-yi* — house-LOC-REL-ACC b. **ev-ni* — house-ACC2

c. *ev-de-ki-ni* — house-LOC-REL-ACC2 d. *ev-i* — house-ACC

6.5.2 Derivations. *-ki* ranges over the case-marked noun, which, as (46a–b) indicate, can be lexical or phrasal. In a lexemic analysis, the entire *ki*-marked noun would have to be rebracketed before the adjective *küçük* can apply to its right scope (which is *ev*, not *çocuk*).

(46) a.

ev *-de* *-ki*
house -LOC -PROki

$$\overset{b}{\triangleleft} N \qquad \overset{c}{\triangleleft} N \backslash \overset{o}{\triangleleft} N \qquad \overset{l}{\bowtie} N \backslash \overset{c}{\bowtie} N_{loc}$$

$$\overline{\overset{c}{\triangleleft} N_{loc}}^{\,<}$$

$$\overline{\overset{l}{\bowtie} N : \lambda f.\mathsf{and}(\mathsf{at\ PRO\ house})(f[\mathsf{PRO}])}^{\,<}$$

'the one that is in the house'

b.

küçük *ev* *-de* *-ki* *çocuk*
little house -LOC -ADJki child

$$\overset{b}{\triangleleft} N / \overset{b}{\triangleleft} N \qquad \overset{b}{\triangleleft} N \qquad \overset{c}{\triangleleft} N \backslash \overset{o}{\triangleleft} N \qquad (\overset{l}{\bowtie} N / \overset{n}{\triangleleft} N) \backslash \overset{c}{\bowtie} N_{loc} \qquad \overset{b}{\triangleleft} N$$

$$\overline{\overset{b}{\triangleleft} N}^{\,>}$$

$$\overline{\overset{c}{\triangleleft} N_{loc}}^{\,<}$$

$$\overline{\overset{l}{\bowtie} N / \overset{n}{\triangleleft} N}^{\,<}$$

$$\overline{\overset{l}{\bowtie} N : \lambda f.\mathsf{and}(\mathsf{at(little\ house)child})(f[\mathsf{child}])}^{\,>}$$

'the child$_i$, the one$_i$ at the little house'

There is another *ki* in Turkish that forms nonrestrictive relative clauses as postmodifiers. It is a Persian borrowing and follows the Indo-European pattern of relative clause formation (47). It can be distinguished from the bound morpheme *-ki* lexically. Its attachment characteristic is also different than that of *-ki* (44e).

(47) *Adam ki hep uyur*
man that always sleep-TENSE
'the man, who always sleeps'

6.6 Possessive Constructions and Syntactic Compounds

The grammatical marking of possession is realized through the genitive case on the possessor (N_{gen}) and the possessive marker on the possessee (N_{poss}). N_{gen} and N_{poss} must agree in person and number (48a), and the resulting noun group is configurational. Possessives can be nested (48c).

(48) a. *ev-in* *kapı-sı*
house-GEN3 door-POSS3s
'the door of the house'

b. * *ev-in kapı* / **ev-in kapı-lar* (door-PLU)

c. *ben-im* *arkadaş-ım-ın* *ev-i-nin* *kapı-lar-ı*
I-GEN1 friend-POSS1s-GEN3 house-POSS3s door-PLU-POSS3s
'my friend's house's doors'

d. *ben-im* *arkadaş-ım-ın*$_i$ *dost-u-nun*$_j$ *kendisi*$_{*i/j}$
I-GEN1 friend-POSS1s-GEN3 buddy-POSS3s-GEN3 self
'my friend's buddy himself'

e. *Her* *çalışan-ın* *bazı* *hak-lar-ı* *vardır*
every worker-GEN3 some right-PLU-POSS3s exists
$\forall x \exists y((worker(x) \wedge right(y)) \rightarrow has(x,y))$
but not $\exists y \forall x(right(y) \wedge (worker(x) \rightarrow has(x,y)))$

6.6.1 Lexical Types for Possessives. Type assignments for the genitive and the possessive can be schematized over person (p) and number (n) features, as in (49).

(49)
$$\text{-GEN}_{pn} \quad := \quad \overset{a}{\circ}\ s - (\overset{o}{\triangleleft} N/(\overset{o}{\bowtie} N_{pn}\backslash \overset{o}{\bowtie} N_{pn}))\backslash \overset{o}{\triangleleft} N_{pn} \colon \lambda x.\lambda y.\mathsf{poss}\, yx$$
$$\text{-POSS}_{pn} \quad := \quad \overset{a}{\circ}\ s - (\overset{o}{\bowtie} N_{pn}\backslash \overset{o}{\bowtie} N_{pn})\backslash \overset{n}{\triangleleft} N_{pn} \colon \lambda f.f$$

The possessive marker's result category is a functor because it enforces agreement with the type raised specifier.[18] (48d–e) indicate that the genitive marker is a type raiser; the possessor scopes over the possessee. For morphosyntactic modality, the genitive marker can be attached to nouns that are inflected up to and including a possessive marker ($\overset{o}{\triangleleft} N$). Moreover, nesting in possessives implies that the specifier may be a genitive. Hence, the stem's category must be $\overset{o}{\triangleleft} N$.

But there is a finer control over the possessee argument's category, because it *must* be inflected with the possessive marker to signify relation of possession (cf. (48a–b)). Semantically, the possessive must outscope nominal modification. For instance, (50a) has the PAS as indicated, hence both markers must range over a noun group, not just

18 An "inert" category such as N may be motivated by the prodrop phenomenon, in which the specifier may be dropped under pragmatically conditioned circumstances. But this analysis disregards the point that binding relations (hence semantics) still require the coindexation of the specifier with some overt referent, which can be inferred from the discourse. Such an interface phenomenon seems to be better suited for handling by interactions in the components of a multidimensional grammar, rather than as a purely syntactic phenomenon.

the stem. Binding relations require an organization of the type (poss possee possessor) (50b–c).

(50) a. *yaşlı adam-ın küçük kız-ı*
old man-GEN3 little daughter-POSS3s
'old man's little daughter' = poss(little daughter)(old man)

b. *adam$_i$-ın kendi$_i$-si*
man-GEN self-POSS
'the man himself'

c. **kendi$_i$ adam$_i$-ı*

6.6.2 Derivation of Possessive Constructions. Example (51) shows the wide scope of the genitive (51a) and nested genitives (51b).

(51) a.

yaşlı	*adam*	*-ın*	*küçük*	*kız*	*-ı*
old	man	-GEN	little	daughter	-POSS
$\overset{b}{\triangleleft} N/\overset{b}{\triangleleft} N$	$\overset{b}{\triangleleft} N$	$\overset{o}{\triangleleft} N/(\overset{o}{\bowtie} N\backslash\overset{o}{\bowtie} N)\backslash\overset{o}{\triangleleft} N$	$\overset{b}{\triangleleft} N/\overset{b}{\triangleleft} N$	$\overset{b}{\triangleleft} N$	$\overset{o}{\bowtie} N\backslash\overset{o}{\bowtie} N\backslash\overset{n}{\triangleleft} N$

yaşlı adam (>): $\overset{b}{\triangleleft} N$; küçük kız (>): $\overset{b}{\triangleleft} N$

yaşlı adam-ın (<): $\overset{o}{\triangleleft} N/(\overset{o}{\bowtie} N\backslash\overset{o}{\bowtie} N)$; küçük kız-ı (<): $\overset{o}{\bowtie} N\backslash\overset{o}{\bowtie} N$

(>): $\overset{o}{\triangleleft} N$: poss(little daughter)(old man)

'old man's little daughter'

b.

ben	*-im*	*arkadaş*	*-ım*	*-ın*	*ev*	*-i*
I	-GEN	friend	-POSS	-GEN	house	-POSS
N	$N/(N\backslash N)\backslash N$	N	$N\backslash N\backslash N$	$N/(N\backslash N)\backslash N$	N	$N\backslash N\backslash N$

ben-im (<): $N/(N\backslash N)$; arkadaş-ım (<): $N\backslash N$; ev-i (<): $N\backslash N$

(>): N

(<): $N/(N\backslash N)$

(>): N: poss house(poss friend i)

'my friend's house'

6.6.3 Lexical Types for Compounds. Syntactic compounds exhibit syntactic patterns similar to possessive constructions, but they signify semantic relations of a different kind. In what follows, we use the function comp to signify that the arguments in the PAS form a compound but say nothing about the range of productivity of this function. The lexical semantics of the arguments and a qualia structure (Pustejovsky 1991) may indicate the function's range of applicability. Lexical type assignments for compound markers are as in (52).

(52) -COMP := $\overset{a}{\circ}\; s - \overset{m}{\bowtie} N \backslash \overset{n}{\lhd} N \backslash \overset{b}{\bowtie} N : \lambda x.\lambda y.\mathsf{comp}\, xy$

-COMP2 (nested comp) := $\overset{a}{\circ}\; s - \overset{m}{\bowtie} N \backslash \overset{m}{\bowtie} N \backslash \overset{n}{\lhd} N \backslash \overset{b}{\bowtie} N : \lambda x.\lambda y.\lambda z.\mathsf{comp}(\mathsf{comp}\, xy)z$

Syntactic compounds are formed by means of compound markers that are attached to the head of the compound. For morphosyntactic modality, nonreferentiality of the head implies no inflection ($\overset{b}{\bowtie} N$) or modification (53a–b). The left component can be a noun group (53c) in which there is ambiguity in the scope of modification. This is regulated by typing, for example, the intersective adjectives ambiguous as noun modifiers ($\overset{n}{\lhd} N / \overset{n}{\lhd} N$) and compound modifiers ($\overset{m}{\bowtie} N / \overset{m}{\bowtie} N$).[19] The overall compound may be inflected only for case (see, e.g., (53d) and (53e)).

(53)

a. *otobüs bilet-i*
bus ticket-COMP
'bus ticket'

b. **otobüs yeşil bilet-i*
green

c. *yeşil otobüs bilet-i*
green bus ticket-COMP
green(comp ticket bus)
or comp(ticket(green bus))

d. *otobüs bilet-i-ni*
ticket-COMP-ACC2

e. **otobüs bilet-i-si*
ticket-COMP-POSS

Compound markers serve the dual function of compounding and agreement in possessive constructions; double marking of the possessive is suppressed (cf. 54a–b). The -COMP2 type assignment in (52) handles nested compounds.

(54)

a. *banka-nın faiz oran-ı*
bank-GEN interest rate-COMP.POSS
'interest rate of the bank'

b. **banka-nın faiz oran-ı-sı*
rate-COMP-POSS

We claim that plural compounds are lexically composite functions in a similar vein. This claim has some empirical support from the lexicalization of *-leri* as a third person plural possessive marker; see (55b–c). It follows that *-leri* has the lexi-

19 I am grateful to the anonymous reviewer who proposed this alternative.

cal types of -COMP and -COMP2 with plural and possessive composition: $\lambda x.\lambda y.\mathsf{plu}\,(\mathsf{comp}\,xy)$.

(55) a. *otobüs bilet-leri*
bus ticket-COMP.PLU
'bus tickets'

b. *onlar-ın ev-leri*
they-GEN3 house-POSS3p
'their house'

c. *onlar-ın ev-ler-i*
they-GEN3 house-PLU-POSS3s
'their houses'

6.6.4 Derivation of Compounds. (56) exemplifies derivations with the type assignments in (52). (56a–b) show that both the narrow and the wide scope of the modifier can be accounted for. (56c–d) show that the compound marker interacts with the possessive. Hence, it must carry both poss and comp in possessive constructions involving compounds. (56e–f) are examples of nested compounds. (56f–g) show the effect of strict control ($\overset{b}{\bowtie}N$) over the compound's head.

(56) a. *yeşil* *otobüs* *bilet* *-i*
green bus ticket -COMP

$\overset{n}{\triangleleft}N/\overset{n}{\triangleleft}N \quad \overset{b}{\triangleleft}N \quad \overset{b}{\triangleleft}N \quad \overset{m}{\bowtie}N\backslash\overset{n}{\triangleleft}N\backslash\overset{b}{\bowtie}N$

$\overset{n}{\triangleleft}N$ (>) $\quad \overset{m}{\bowtie}N\backslash\overset{n}{\triangleleft}N$ (<)

$\overset{m}{\bowtie}N: \mathsf{comp\ ticket(green\ bus)}$ (<)

b. *yeşil* *otobüs* *bilet* *-i*

$\overset{m}{\bowtie}N/\overset{m}{\bowtie}N \quad \overset{b}{\triangleleft}N \quad \overset{m}{\bowtie}N\backslash\overset{n}{\triangleleft}N$ (<)

$\overset{m}{\bowtie}N$ (<)

$\overset{m}{\bowtie}N: \mathsf{green(comp\ ticket\ bus)}$ (>)

c. *banka* *-nın* *faiz* *oran* *-ı*
bank -GEN interest rate -COMP.POSS

$N \quad N/(N\backslash N)\backslash N \quad N \quad N \quad N\backslash N\backslash N\backslash N$

$N/(N\backslash N)$ (>) $\quad N\backslash N\backslash N$ (<)

$N\backslash N$ (<)

$N: \mathsf{poss(comp\ rate\ interest)bank}$ (>)
'interest rate of the bank'

d.

banka	*-nın*	*faiz*	*oran*	*-ları*
bank	-GEN	interest	rate	-COMP.POSS.PLU
N	$N/(N\backslash N)\backslash N$	N	N	$N\backslash N\backslash N\backslash N$
$N/(N\backslash N)$ (>)			$N\backslash N\backslash N$ (<)	
		$N\backslash N$ (<)		
N: poss(plu(comp rate interest))bank (>)				

'interest rates of the bank'

e.

kredi	*kart*	*-ı*	*faiz*	*oran*	*-ı*
credit	card	-COMP	interest	rate	-COMP2
N	N	$N\backslash N\backslash N$	N	N	$N\backslash N\backslash N\backslash N$
	$N\backslash N$ (<)			$N\backslash N\backslash N$ (<)	
N (<)			$N\backslash N$ (<)		
N: comp(comp rate interest)(comp card credit) (<)					

'credit card interest rate'

f. *kredi kart-ı yıllık faiz oran-ı*
annual
'credit card annual interest rate'

g. **kredi kart-ı faiz yıllık oran-ı*

7. Conclusion

Theoretical and computational commitment to word-based grammar—and to regard inflectional morphology as a word-internal process—puts artificial limits on specifying the syntactic and semantic domains of all meaning-bearing elements and on the transparent projection of scope from the lexicon. Designating words as minimal units of the lexicon is too constraining for many languages. This traditional notion is also challenged in current linguistic theorizing (e.g., Jackendoff 1997 and Keenan and Stabler 1997). Marslen-Wilson (1999) argues on psycholinguistic grounds that the lexicon must be morphemic even for morphologically simpler languages such as English.

We have argued in this article that the key to the integration of inflectional morphology and syntax is granting representational status to morphemes, which, in a computational system, requires certain precautions. What we propose is enriching the expressive power of the combinatory morphemic lexicon to factor in morphosyntactic types and attachment modalities. Coupled with flexible constituency in the grammar and directionality information coming from the lexicon, these extensions provide the grammar with the information it requires to compute the transparent semantics of morphosyntactic phenomena. This flexibility causes neither inefficiency in parsing nor uncontrolled expressivity. The extensions do not affect the polynomial worst-case complexity results, and category unity is preserved by lattice consistency. The result is a morphemic grammar–lexicon with computationally desirable features such as modularity and transparency. The system is available at ftp://ftp.lcsl.metu.edu.tr/pub/tools/msccg.

Acknowledgments

I am very grateful to four anonymous *CL* reviewers for extensive commentary and suggestions. Thanks to Wolf König and Stefan Müller for the data; and to Jason Baldridge, Gann Bierner, Aysenur Birtürk, Ruken Çakici, Nissim Francez, Stasinos Konstantopoulos, Markus Kracht, Geert-Jan Kruijff, Alan Libert, Mark Steedman, Ümit Turan and Deniz Zeyrek for comments, advice, and criticism.

References

Aho, Alfred V. and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation, and Compiling*. Vol. 1. Prentice-Hall.

Anderson, Stephen R. 1982. Where's morphology? *Linguistic Inquiry*, 13(4):571–612.

Bach, Emmon. 1983. On the relationship between word-grammar and phrase-grammar. *Natural Language and Linguistic Theory*, 1:65–89.

Baldridge, Jason. 1999. Strong equivalence of CCG and Set-CCG. Unpublished manuscript, University of Edinburgh.

Bar-Hillel, Yehoshua, Chaim Gaifman, and Eliyahu Shamir. 1960. On categorial and phrase structure grammars. *Bulletin of the Research Council of Israel*, 9F:1–16.

Bozsahin, Cem. 1998. Deriving the Predicate-Argument structure for a free word order language. In *Proceedings of COLING-ACL 1998*, Montreal, pages 167–173.

Bozsahin, Cem. 2000a. Directionality and the lexicon: Evidence from gapping. Unpublished manuscript, Middle East Technical University, Ankara.

Bozsahin, Cem. 2000b. Gapping and word order in Turkish. In *Proceedings of the 10th International Conference on Turkish Linguistics*, Istanbul.

Bozsahin, Cem and Elvan Göçmen. 1995. A categorial framework for composition in multiple linguistic domains. In *Proceedings of the 4th International Conference on Cognitive Science of NLP*, Dublin.

Bresnan, Joan. 1995. Lexical–Functional syntax. Course notes. Seventh European Summer School in Logic, Language, and Information, Barcelona.

Calder, Jonathan, Ewan Klein, and Henk Zeevat. 1988. Unification categorial grammar. In *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, pages 83–86.

Carpenter, Bob. 1992. Categorial Grammar, lexical rules, and the English predicative. In R. Levine, editor, *Formal Grammar: Theory and Application*. Oxford University Press, pages 168–242.

Carpenter, Bob. 1997. *Type-Logical Semantics*. MIT Press.

Carpenter, Bob. 1999. The Turing-completeness of Multimodal Categorial Grammars. In *Papers Presented to Johan van Benthem in Honor of his 50th Birthday*. ESSLLI, Utrecht.

Carpenter, Bob and Gerald Penn. 1994. *The Attribute Logic Engine User's Guide, Version 2.0*. Carnegie Mellon University.

Chomsky, Noam. 1970. Remarks on nominalization. In R. Jacobs and P. Rosenbaum, editors, *Readings in English Transformational Grammar*. Ginn, Waltham, MA, pages 184–221.

Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press.

Collins, Michael. 1997. Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the ACL*.

Creider, Chet, Jorge Hankamer, and Derick Wood. 1995. Preset two-head automata and natural language morphology. *International Journal of Computer Mathematics*, 58:1–18.

Dalrymple, Mary and Ronald M. Kaplan. 2000. Feature indeterminacy and feature resolution. *Language*, 76:759–798.

Dowty, David. 1979. *Word Meaning and Montague Grammar*. Kluwer, Dordrecht.

Dowty, David. 1991. Toward a minimalist theory of syntactic structure. In *Tilburg Conference on Discontinuous Constituency*, January 1989.

Dowty, David. 1996. Non-constituent coordination, wrapping, and Multimodal Categorial Grammars. In *International Congress of Logic, Methodology, and Philosophy*, Florence, August.

Eisner, Jason. 1996. Efficient normal-form parsing for Combinatory Categorial Grammar. In *Proceedings of the 34th Annual Meeting of the ACL*, pages 79–86.

Fong, Sandiway. 1991. *Computational Properties of Principle-Based Grammatical Theories*. Ph.D. dissertation, MIT.

Güngördü, Zelal and Kemal Oflazer. 1995. Parsing Turkish using the Lexical-Functional Grammar formalism. *Machine Translation*, 10:293–319.

Hankamer, Jorge. 1989. Morphological parsing and the lexicon. In W. Marslen-Wilson, editor, *Lexical Representation and Process*. MIT Press.

Hepple, Mark. 1990a. *The Grammar and Processing of Order and Dependency: A Categorial Approach*. Ph.D. dissertation, University of Edinburgh.

Hepple, Mark. 1990b. Normal form theorem proving for the Lambek Calculus. In *Proceedings of COLING 1990*.

Hepple, Mark and Glyn Morrill. 1989. Parsing and derivational equivalence. In *Proceedings of the 4th EACL*, Manchester.

Heylen, Dirk. 1997. Underspecification in Type-Logical Grammars. In *Logical Aspects of Computational Linguistics (LACL)*, Nancy.

Heylen, Dirk. 1999. *Types and Sorts: Resource Logic for Feature Checking*. Ph.D. dissertation, Utrecht University.

Hockenmaier, Julia, Gann Bierner, and Jason Baldridge. 2000. Providing robustness for a CCG system. Unpublished manuscript, University of Edinburgh.

Hoeksema, Jack. 1985. *Categorial Morphology*. Garland, New York.

Hoeksema, Jack and Richard D. Janda. 1988. Implications of process-morphology for Categorial Grammar. In Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler, editors, *Categorial Grammars and Natural Language Structures*. D. Reidel, Dordrecht, pages 199–247.

Hoffman, Beryl. 1995. *The Computational Analysis of the Syntax and Interpretation of "Free" Word Order in Turkish*. Ph.D. dissertation, University of Pennsylvania.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. MIT Press.

Jacobson, Pauline. 1996. The syntax/semantics interface in Categorial Grammar. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*. Blackwell, 89–116.

Janeway, Roger. 1990. Unacceptable ambiguity in Categorial Grammar. In *Proceedings of the Ninth West Coast Conference on Formal Linguistics*, pages 305–316.

Johnson, Mark. 1988. Deductive parsing with multiple levels of representation. In *Proceedings of the 26th Annual Meeting of the ACL*, pages 241–248.

Johnson, Mark and Sam Bayer. 1995. Features and agreement in Lambek Categorial Grammar. In *Proceedings of the 1995 ESSLLI Formal Grammar Workshop*, pages 123–137.

Jurafsky, Daniel and James H. Martin. 2000. *Speech and Language Processing*. Prentice-Hall.

Karttunen, Lauri. 1989. Radical lexicalism. In Mark Baltin and Anthony Kroch, editors, *Alternative Conceptions of Phrase Structure*. University of Chicago Press, pages 43–65.

Keenan, Edward L. and Edward Stabler. 1997. Bare grammar. Course notes, Ninth European Summer School on Logic, Language, and Information, Aix-en-Provence.

Komagata, Nobo. 1997. Efficient parsing for CCGs with generalized type-raised categories. In *Proceedings of the Fifth Int. Workshop on Parsing Technologies*, pages 135–146.

König, Esther. 1989. Parsing as natural deduction. In *Proceedings of the 27th Annual Meeting of the ACL*, pages 272–279.

Kracht, Markus. 1999. Referent systems, argument structure, and syntax. ESSLLI lecture notes, Utrecht.

Kruijff, Geert-Jan M. and Jason M. Baldridge. 2000. Relating categorial type logics and CCG through simulation. Unpublished manuscript, University of Edinburgh.

Lambek, Joachim. 1958. The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–170.

Marslen-Wilson, William. 1999. Abstractness and combination: The morphemic lexicon. In Simon Garrod and Martin J. Pickering, editors, *Language Processing*. Psychology Press, East Sussex, UK, pages 101–119.

Miller, Philip H. and Ivan A. Sag. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory*, 15:573–639.

Moortgat, Michael. 1988a. *Categorial Investigations: Logical and Linguistic Aspects of the Lambek Calculus*. Foris, Dordrecht.

Moortgat, Michael. 1988b. Mixed composition and discontinuous dependencies. In Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler, editors, *Categorial Grammars and Natural Language Structures*. D. Reidel, Dordrecht, pages 319–348.

Moortgat, Michael and Richard T. Oehrle. 1994. Adjacency, dependency and order. In *Proceedings of the Ninth Amsterdam Colloquium*.

Morrill, Glyn V. 1994. *Type Logical Grammar: Categorial Logic of Signs*. Kluwer.

Morrill, Glyn V. 1999. Geometry of lexico-syntactic interaction. In *Proceedings of the Ninth EACL*, Bergen.

Müller, Stefan. 1999. *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. Linguistische Arbeiten 394. Max Niemeyer Verlag, Tübingen.

Oflazer, Kemal. 1994. Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9(2).

Oflazer, Kemal. 1996. Error-tolerant finite-state recognition with applications to morphological analysis and spelling correction. *Computational Linguistics*, 22:73–89.

Oflazer, Kemal, Elvan Göçmen, and Cem Bozsahin. 1994. An outline of Turkish morphology. Report to NATO Science Division SfS III (TU-LANGUAGE), Brussels.

Oflazer, Kemal, Bilge Say, Dilek Zeynep Hakkani-Tür, and Gökhan Tür. 2001. Building a Turkish treebank. In Anne Abeille, editor, *Building and Exploiting Syntactically-Annotated Corpora*. Kluwer.

Pareschi, Remo and Mark Steedman. 1987. A lazy way to chart-parse with Categorial Grammars. In *Proceedings of the 25th Annual Meeting of the ACL*, pages 81–88.

Pereira, Fernando C. N. and Stuart M. Shieber. 1987. *Prolog and Natural-Language Analysis*. CSLI, Stanford, CA.

Pollard, Carl and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press.

Pulman, Stephen G. 1996. Unification encodings of grammatical notations. *Computational Linguistics*, 22:295–327.

Pustejovsky, James. 1991. The generative lexicon. *Computational Linguistics*, 17(4):409–441.

Sehitoglu, Onur. 1996. A sign-based phrase structure grammar for Turkish. Master's thesis, Middle East Technical University.

Sehitoglu, Onur and Cem Bozsahin. 1999. Lexical rules and lexical organization: Productivity in the lexicon. In Evelyne Viegas, editor, *Breadth and Depth of Semantic Lexicons*. Kluwer, pages 39–57.

Steedman, Mark. 1985. Dependency and coördination in the grammar of Dutch and English. *Language*, 61(3):523–568.

Steedman, Mark. 1987. Combinatory grammars and parasitic gaps. *Natural Language and Linguistic Theory*, 5:403–439.

Steedman, Mark. 1988. Combinators and grammars. In Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler, editors, *Categorial Grammars and Natural Language Structures*. D. Reidel, Dordrecht, pages 417–442.

Steedman, Mark. 1991a. Structure and intonation. *Language*, 67:260–298.

Steedman, Mark. 1991b. Type raising and directionality in Combinatory Grammar. In *Proceedings of the 29th Annual Meeting of the ACL*, pages 71–78.

Steedman, Mark. 1996. *Surface Structure and Interpretation*. MIT Press.

Steedman, Mark. 2000a. *The Syntactic Process*. MIT Press.

Steedman, Mark. 2000b. Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 31(4): 649–689.

Szabolcsi, Anna. 1983. ECP in categorial grammar. Unpublished manuscript, Max-Planck Institute.

Szabolcsi, Anna. 1987. Bound variables in syntax: Are there any? In *Proceedings of the 6th Amsterdam Colloquium*, pages 331–350.

Tomita, Masaru. 1988. *The Generalized LR Parser/Compiler*. Technical report, Center for Machine Translation, Carnegie Mellon University.

Vijay-Shanker, K. and David J. Weir. 1993. Parsing some constrained grammar formalisms. *Computational Linguistics*, 19:591–636.

Wechsler, Stephen and Larisa Zlatić. 2000. A theory of agreement and its application to Serbo-Croatian. *Language*, 76:799–832.

Whitelock, Pete J. 1988. A feature-based categorial morpho-syntax for Japanese. In Uwe Reyle and C. Rohrer, editors, *Natural Language Parsing and Linguistic Theories*. D. Reidel, pages 230–261.

Williams, Edwin. 1981. On the notions "lexically related" and "head of a word." *Linguistic Inquiry*, 12(2):245–274.

Wittenburg, Kent. 1987. Predictive combinators. In *Proceedings of the 25th Annual Meeting of the ACL*, pages 73–79.