LEXICAL RULES AND LEXICAL ORGANIZATION: PRODUCTIVITY IN THE LEXICON

ONUR T. SEHITOGLU AND CEM BOZSAHIN
Department of Computer Engineering
Middle East Technical University
06531, Ankara Turkey

Abstract.

In this paper we outline a lexical organization for Turkish that makes use of lexical rules for inflections, semantically predictable derivations, and lexical category changes to control the proliferation of lexical entries. Lexical rules handle changes in grammatical roles, enforce type constraints, and control the mapping of subcategorization frames in valency-changing operations. A lexical inheritance hierarchy facilitates the enforcement of type constraints. The resulting grammar architecture has morphology embedded in the lexicon.

The design has been tested as part of an HPSG grammar for Turkish. Rebracketing in compounds and phrasal scope in affixation present formidable difficulties for lexical rules. In terms of performance, run-time execution of the rules seems to be a far better alternative than precompilation. The latter causes exponential growth in the lexicon due to intensive use of inflections and derivations in Turkish.

1. Introduction

We report on an experiment to model agglutinating morphology of Turkish in the lexicon by lexical rules. Our motivation was to see the extent to which the lexical rules can act as (a) the interface between inflectional morphology and syntax, (b) the mechanism for handling semiproductive parts of derivational morphology, and (c) the source of reducing (or eliminating) the proliferation of lexical items. Turkish demonstrates a strong interplay of morphology and syntax, which makes morphology a more active component of the grammar. Prominence of morphology puts a greater demand

on the lexical information as well, to the extent that lexical organization in a computational model must take this type of information into account in design. The lexicon may grow to an unmanageable size due to heavy use of inflections and derivations. For instance, the nominal paradigm has four affixes (number, case, possessive, relativizer), and the verbal paradigm has eight (for voice, tense, person, aspect, and mood). Generating the full paradigm for a nominal (verbal) root requires 2^4 (2^8) entries in the lexicon. The problem is further complicated by the rich inventory of derivational affixes for both paradigms, as exemplified in (1).

(1) *Yaz-ıcı-lar-a* *gör-ev-ler-i* *bil-dir-il-me- miş-ti*
write-VtoN-PLU-DAT able-VtoN-PLU-ACC know-CAUS-PASS-NEG-ASP-TENSE
'The clerks have not been informed of their duties'

Handling inflections and derivations with lexical rules necessitates encoding of grammatical changes coming from morphology in the lexicon as well. For instance, a causative suffix will demote an agent to a patient or a recipient, and it will add a new grammatical role for the causer (the new agent). A locative case suffix will mark an NP as an adjunct, which can no longer satisfy subcategorization requirements of the verbs or postpositions. But lexical rules are put to the test in rebracketing and in semantic composition. We elaborate on the consequences of these phenomena in section 3.

Another source for economy of representation can be seen in example (2), where attributive adjectives (2a and 2c) are used as nouns in (2b) and (2d). One solution to this problem is underspecification, i.e., grouping the nouns and adjectives under a single category.[1] An alternative is to use a lexical rule for differentiating the predicate and term readings of the lexical entry. This seems to be a better alternative for unifying the treatments of nouns, adjectives, derived nouns and derived adjectives. For instance, *yaşlı* in (2c) is a derived adjective produced by a lexical rule. The output can be fed into another lexical rule to obtain a zero derivation of a noun from the adjective (2d).

The dynamic nature of lexical rules allows access to lexical semantics on demand, hence giving control over their application. For example, for the derived adjective *güdüm-lü*[2] (lead-ADJ "ballistic"), one would not want to obtain its nominal reading; it is unlikely to be used as a noun. It would require a complex (and perhaps extendible) inheritance hierarchy to achieve the same effect by underspecification.

[1] In fact, traditional Turkish grammar books such as (Lewis, 1967) collectively call them "substantives."

[2] *-lü* is an allomorph of the morpheme that can be represented as *-lI*. Other allomorphs are *-lı* (as in example 2c–d), *-lu* and *-li*.

(2) a. *kuru yaprak*
dry leaf
'dry leaf'

b. *meyve kuru-su*
fruit dry-COMP
'dried fruit'

c. *yaş-lı hanım*
age-ADJ lady
'old lady'

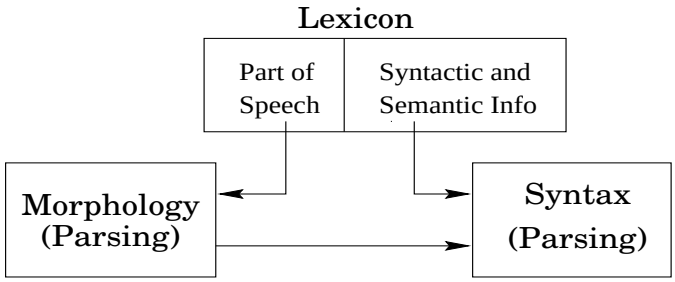
d. *bütün yaş-lı-lar*
all age-ADJ-PLU
'all elderly'

It appears that an agglutinating language will benefit from the lexical rule approach to morphology. Lexical rules can make use of lexical information to model morphotactic, syntactic, and semantic aspects of morphology, thereby enriching the lexicon without an exponential growth. Lexical rules for inflections contribute to grammatical meaning composition, and lexical rules for derivations provide a productive way of word formation, i.e., composition of lexical meanings. However, most types of lexical rules for morphology will overgenerate if the language is not purely polysynthetic. The information source for constraining the application of the rules is, in most cases, lexical semantics. In what follows, we describe different kinds of lexical rules for the morphology-lexicon-syntax interface. We also discuss processing issues such as run-time generation versus precompiling of word forms.

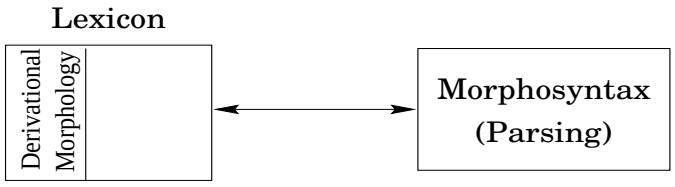
2. Morphology-syntax Interface

Modelling inflections, derivations, and the corresponding phonological alternations via lexical rules places morphology in the lexicon (cf. Figure 1c). Morphemes have no representational status. This is essentially word-based morphology (Aronoff, 1976; Anderson, 1992) by lexical rules. Alternatives to this approach (for Turkish) have also been explored, such as the modularization of syntax and morphology (cf. Figure 1a) which keeps them—and their lexicons—as separate systems (Güngördü and Oflazer, 1995). The flow of information is unidirectional in this architecture. The other alternative is the integrative approach to morphology and syntax, which treats morphosyntax in the same manner as syntax with respect to

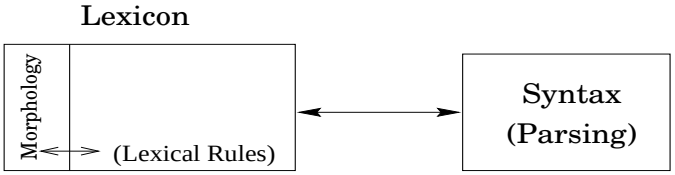
semantic composition[3] (Bozsahin and Göçmen, 1995).



(a) No embedding of morphology in the lexicon (modular).



(b) Partial embedding (integrative).



(c) Full embedding (lexical).

Figure 1. Architectures for Morphology-Syntax Interface.

From a computational point of view, the modular approach has efficient lexical access since lexical search is performed on root forms and bound morphemes are not considered lexical items. This approach tends to isolate the morphological information in the lexicon from syntactic information. This seems to be done out of practical considerations rather than methodological necessity; a finite-state component takes care of morphotactics with minimal use of lexical information (e.g., lexical category of the root forms), while a phrase structure component takes care of syntax. This separation, however, causes overgeneration in derivational morphology since lexical semantic information which could rule out exceptional cases in

[3]This approach is integrative in the sense that the lexicon contains bound forms as well as free forms, and derivational morphology is considered as part of the lexicon (cf. Figure 1b). For it to be fully integrated, the derivational affixes need to have a lexical representation as well. This is akin to morpheme-based morphology of Lieber (1992).

semiproductive derivational affixes will be inaccessible to the morphological component. For instance, the adverbial affix *-leyin* attaches to temporal nouns (e.g., *gündüz* "day", *gece* "night") to produce temporal adverbs (*gündüzleyin, geceleyin* "during the day/night"). It would overgenerate if it is allowed to attach to any noun.

Causativization is also a good example for morphology-syntax-semantics interaction. It is very productive; any intransitive (transitive) verb can be causativized into a transitive (ditransitive) verb. But some causative forms have been lexicalized. For instance, *git* ("go") has the lexicalized causative *götür* ("deliver"). To block *git*-CAUSE, the morphological component must have access to semantic information (e.g., [∓causative]) about the lexical item. Causatives also introduce semantic constraints on the predicate-argument structure of the causativized predicate. For instance, the derived verb *aç-tır* (open-CAUS) requires a causee, which is marked dative. The causee may be inferred from the context as well. But if there is a dative NP which cannot fulfill the argument requirements (e.g., [-animate]), an adverbial reading is forced (cf. 3a and 3b, from (Göksel, 1993, p.106–108)).

(3) a. *Nuran Deniz'e kapı-yı aç-tır-dı*
N.NOM D-DAT door-ACC open-CAUS-TENSE.AGR
'Nuran made Deniz open the door.'

b. *Nuran balkon-a kapı-yı aç-tır-dı*
balcony-DAT
'Nuran made (someone) open the door to the balcony.'

In the integrative approach, the lexicon contains free and bound morphemes; they have complete syntactic and semantic specifications. As lexical items, they must take part in semantic composition which goes on in parallel with morphological and syntactic processing. Some of the inflections, e.g. person, perform only feature marking, hence their semantic form is that of identity. Grammatical cases such as the accusative affect the LF by filling argument positions. Concrete cases such as the locative introduce an adverbial (e.g., $(at\ SF_s)$ where SF_s is the semantic form of the stem to which the case marker is attached). The attachment takes place in the grammar, thus the grammar must differentiate between morphological and syntactic composition to eliminate overgeneration. For instance, it must rule out affixation of two words or concatenation of two bound morphemes (if the latter is disallowed in the language).

With respect to the place of morphology in the grammar architecture, the lexical approach represents the other extreme of the modular approach. Morphology is a subcomponent of the lexicon. However, as far as the complexity and coverage of the lexicon are concerned, it is in the midpoint of the

three approaches. In this view, morphology is not isolated from syntax, but, similar to the modular organization, bound morphemes are not considered lexical items. They are attached to stems via lexical rules. This implies that lexical rules are responsible for semantic composition and for the changes in syntactic requirements. The integrative approach puts a greater burden on syntax and the lexicon because of their greater dependence on morphological information. Syntax must be sensitive to the different attachment characteristics of the lexemes and the morphemes (affixes, clitics), and the lexicon must contain information about the bound morphemes.

The degree of integration of morphology and syntax is a question of considerable debate in linguistic theory. GB and LFG explicitly state some degree of integration by considering inflectional morphology as part of syntax, and treating derivational morphology as a matter of the lexicon. Dowty (1979) describes a lexical theory with syntactic rules and lexical rules cutting across the morphology-syntax boundary. Inflectional morphology and semantically predictable parts of derivational morphology are modelled by syntactic rules for morphology. Semiproductive and semantically less predictable parts of derivational morphology are handled by lexical rules for morphology. Thus, semantic predictability is a decisive issue in integration.

Besides its implications for universal grammar, keeping morphology and syntax entirely separate forces one to stipulate different scopes for affixes. For instance, the adverbial suffix *-ken* and the adjectival *-lu* might have phrasal (4a and 4c) or lexical scope (4b and 4d). The bracketing in (4a) and (4b) can be disambiguated by looking at the valence of the verbal stem for *-ken* (transitive in the first case). The integrative approach allows affixes to 'pick out' different scopes in mixed morphological and syntactic composition. The lexical approach can accommodate both readings provided that the lexical rules can make use of *local* syntactic information available in the lexical item (e.g., valence). Morphologically ambiguous cases such as (5a–c) are handled by multiple instantiations of the lexical rules.

(4) a. *Çocuk kutu-yu [kız-a bakar]-ken açtı*
child box-DAT girl-DAT look-ADV open
'The child opened the box while watching the girl.'

b. *Çocuklar [yürür]-ken taş toplamışlar*
children walk-ADV stone picked
'The children have picked stones while walking.'

c. *[Uzun kol]-lu gömlek*
long sleeve-ADJ shirt
'shirt with long sleeves'

d. *Uzun* [*çiçek*]-*li* *gömlek*
long flower-ADJ shirt
'long shirt with flower patterns'

(5) a. *kalem-ler-i*
pencil-PLU-ACC
'the pencils (=object)'

b. *kalem-ler-i*
pencil-PLU-POSS.3SG
'his/her pencils'

c. *kalem-leri*
pencil-POSS.3PL
'their pencils'

It is too early to evaluate the advantages and disadvantages of these approaches in terms of competence grammars and performance issues. But the choice of the strategy also affects the design of lexical organization. For instance, if inflections and derivations are handled by lexical rules, the morphological information need not be stored at all, since the lexical rules will reflect the changes in syntactic and semantic requirements coming from morphology. If morphology is treated almost like syntax, lexical knowledge should contain richer morphological information, including a semantic representation for bound forms, information about boundedness/freeness of morphemes, and the type of attachment (e.g., affixation, cliticization, syntactic concatenation) (Bozsahin and Göçmen, 1995; Hoeksema and Janda, 1988).

In this study, a lexical inheritance hierarchy is used in conjunction with the lexical rules. It allows a regimented representation for feature structures, and enforces type constraints on free forms (words). Bound forms are not part of the lexicon. The hierarchy is given in Figure 2.

This tree is part of a greater hierarchy which includes inheritance information for words and phrases. We make use of the inheritance and type checking mechanism of ALE (Carpenter and Penn, 1994) to impose type specific constraints on words. Words are distinguished from phrases by disallowing any kind of gapping below the word level in the tree. Designating a lexical item as one of the subtypes in the hierarchy will apply all the constraints and incorporate the feature structures of the supertypes along the path to *word*. For instance, a qualitative adjective (e.g., *rahat* "comfortable") is distinguished from a quantitative one (e.g., *çift* "double") by its choice of modifiers; the latter does not allow intensifiers (6).

(6) a. *çok* *rahat* *koltuk*
very comfortable couch
'very comfortable couch'

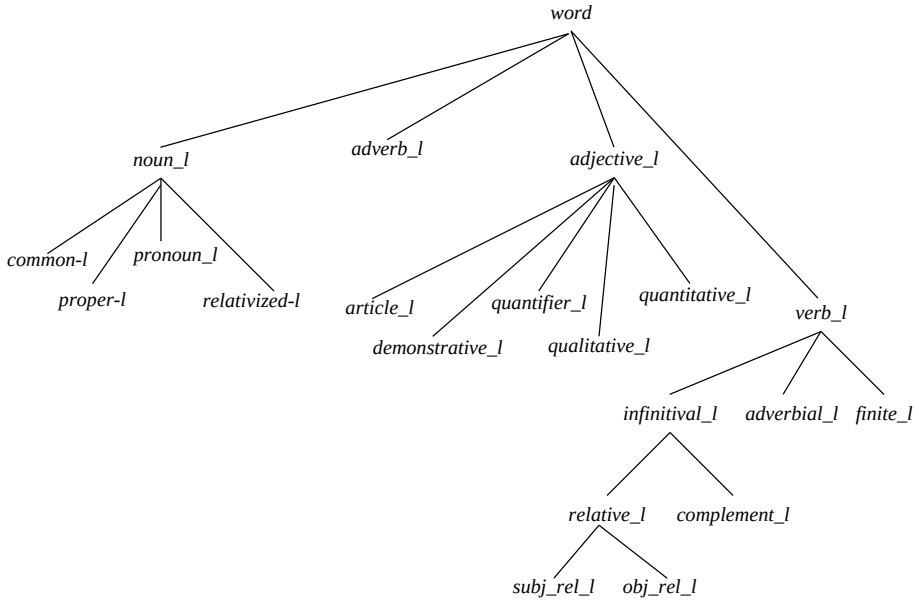


Figure 2. Lexical hierarchy

b. * *çok çift koltuk*
c. *rahat çift koltuk*
comfortable double couch
'comfortable twin couch'

The fragments of the type constraints[4] for these subtypes are given in Figure 3. The controlled use of type constraints at different levels of the lexical hierarchy eliminate the need to enumerate type specific lexical rules to achieve the same effect.

3. Types of lexical rules

We divide the lexical rules for morphology into three main groups. The first group is for inflectional morphology, which marks the lexical item with features required by syntax, e.g., person, number, and case. The second group of lexical rules models the changes in the syntactic frame of the lexical item. These include grammatical function changing affixes (reflexive, causative, reciprocal, passive), infinitive markers for subordination, and some 'zero' inflections such as nonovert marking of the indefinite object. The third group

[4]We use HPSG style feature structures and signatures in our descriptions. See Pollard and Sag (1994).

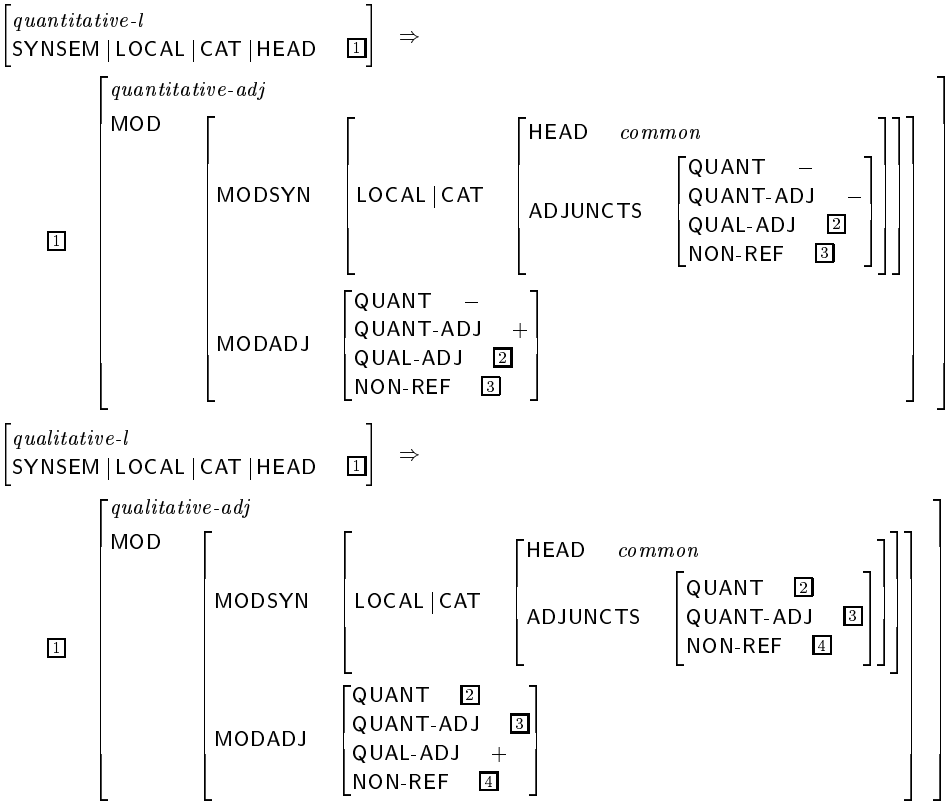


Figure 3. Type constraints for words and some subtypes.

is for the category-changing operations, i.e. for derivational morphology. We provide examples of these rules below.

There is no separate morphological parsing component in this approach; the lexical rules incorporate a model of morphotactics for proper ordering of the affixes. This is achieved by the sign constraints on the lexical items. Since the feature structure of a lexical item includes all information about its phonological, syntactic, and semantic properties, restrictions on all these aspects can be put in a single constraint. Morphology-phonology interface is handled in a constrained-based fashion as well. Phonological alternations in stems and affixes (e.g., vowel harmony, devoicing, epenthesis etc.) are specified as constraints on the phonological sign of the stem. The applicability of the lexical rule for an affix is determined by the phonological constraints on the input and the phonological properties of the allomorphs. For instance, the locative case marker has allomorphs *-de/-da/-te/-ta*. Vowel harmony and voicing constraints determine their proper phonological context during

morphological composition. These kinds of rules are not lexical rules per se since they do not operate on lexical properties of the words. Nevertheless, they are necessary for obtaining the right surface word forms in parsing or generation.

3.1. FEATURE MARKING OPERATIONS

These rules reflect the grammatical requirements imposed by inflections. Overt case markers that can encode arguments (e.g., accusative, dative, genitive) only affect the CASE feature of the syntactic frame. Adjunct markers must make additional changes. For instance, the locative case suffix in Turkish also marks an NP as adjunct (7).

(7) *Adam araba-da uyu-du* .
man car-LOC sleep-TENSE.AGR
'The man slept in the car'

The lexical rule for the locative case is given in Figure 4. This rule is applied when the locative suffix is attached to a nominal stem. In HPSG terms, the head of the NP is marked with the locative case, and the type of NP is changed to an adjunct. This is achieved by modifying the head feature MOD: While the nominative-marked noun has null value, a MODSYN value with verbal head is introduced in the head feature of the locative noun. This will allow the locative-marked noun to modify a verb. It cannot be an argument of a verb or a postposition. This issue is critical for parsing relatively free word order languages where grammatical functions are often indicated by overt case marking rather than by configurational means. Figure 4 also shows the derivation of the semantic representation for the case-marked NP; *at(x,y)* is a second-order predicate that holds between a term y and a predicate x. This predicate is inserted into the set of restrictions for the noun. For (7), this method produces *at(sleep(man),car)*.

3.2. FRAME CHANGING OPERATIONS

There are some phenomena in Turkish that are on the boundary of lexicon and syntax, which we opted to model in the lexicon, e.g., nonreferential objects, and valency change in the causatives. In the following, we briefly describe the lexical rules for them.

Case marking is overt in Turkish, which allows for scrambling of the constituents. All six permutations of the SOV order are felicitous if the object NP is case marked (e.g., 8a and 8b). If the object is nonreferential or indefinite (cf. 8a and 8c), it is not marked morphologically, which blocks scrambling, and the unmarked SOV order is used (cf. 8c and 8d).

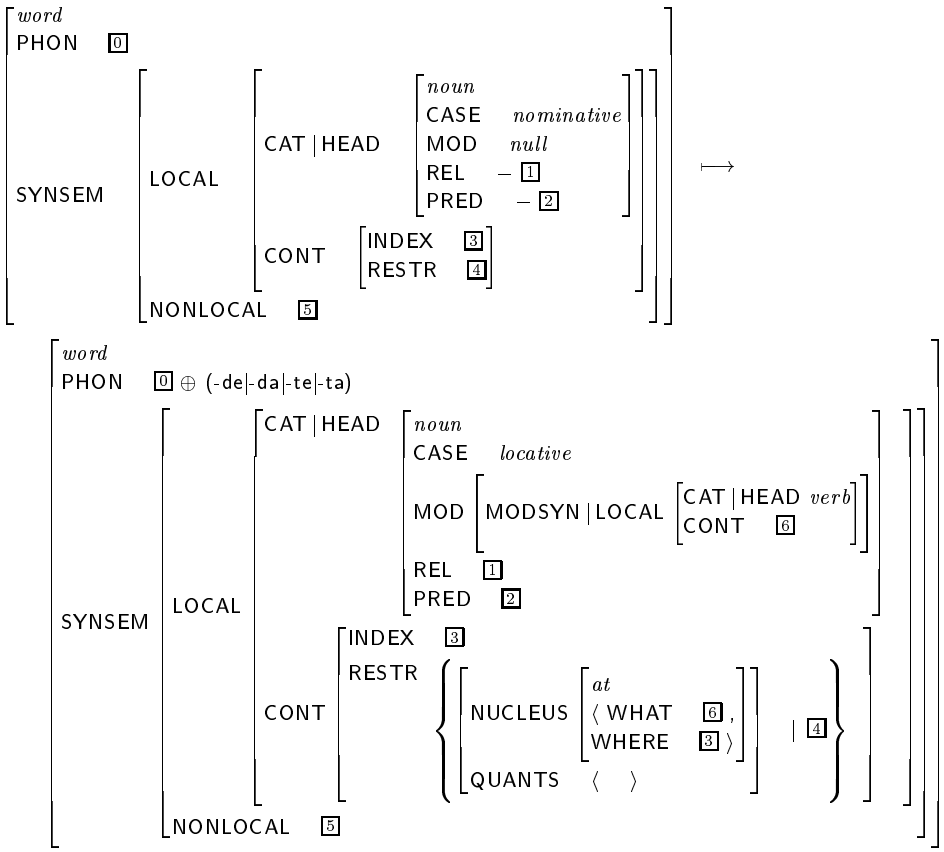
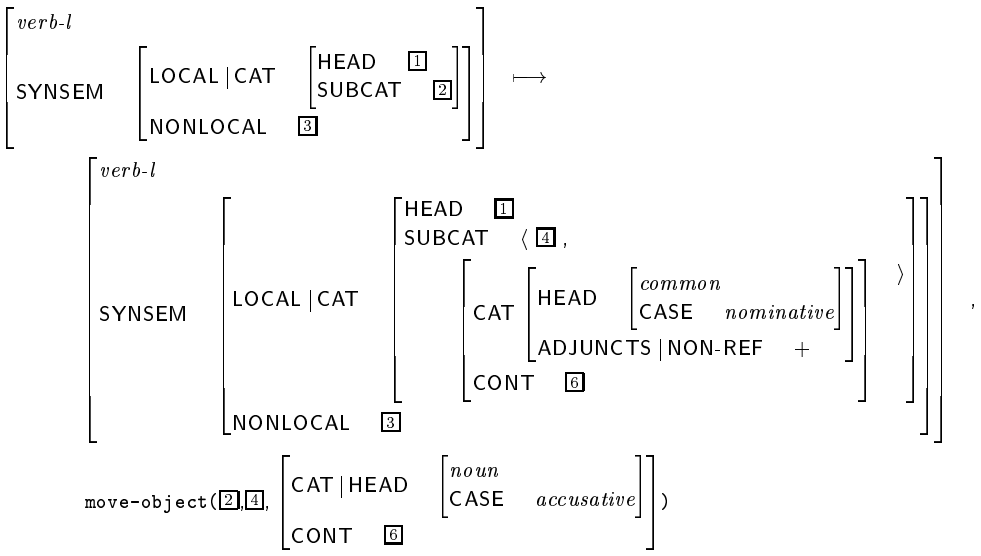


Figure 4. The lexical rule for the locative case.

(8) a. *Çocuk kitab-ı oku-du*
child.NOM book-ACC(=object) read-TENSE.AGR
'The child read the book.'
b. *Kitab-ı çocuk oku-du*
c. *Çocuk kitap oku-du*
child.NOM book.ACC(=object) read-TENSE.AGR
'the child read a book (≅ the child did book-reading)'
d. * *Kitap çocuk okudu*

Nonreferential objects are not morphologically marked for case, and they must occupy the immediately preverbal position. One way of dealing with nouns, then, is to keep two entries in the lexicon: one for the unmarked form which may receive case marking and scramble, and one with lexically assigned accusative case, which may not scramble. Our solution is to have

a lexical rule that changes the subcategorization frames of verbs to handle cases where objects may be case-marked NPs or unmarked Ns. In the second case, the entity is marked indefinite and all scrambling is blocked by the lexical rule. Figure 5 shows the rule in HPSG notation (the rule is simplified for ease of exposition).



Where `move-object` is a definite clause which deletes the accusative object from the SUBCAT structure in first argument and return resulting structure and accusative object in second and third argument respectively.

Figure 5. Lexical rule for nonreferential objects.

Morphological causatives are modelled in a similar vein. A causative suffix changes the subcategorization frame of the verb by adding one more argument and changing the grammatical constraints on the other arguments. For instance, the new argument becomes the subject (causer), and the old subject (agent) is demoted down the grammatical hierarchy (Comrie, 1976) to direct object or indirect object, depending on the argument structure:

(9) a. *Can arkadaş-ı-nı çağır-dı*
friend-POSS-ACC call-TENSE.AGR
'Can called his friend.'

b. *Mehmet Can-a arkadaş-ı-nı çağır-t-tı*
Can-DAT friend-POSS-ACC call-CAUS-TENSE.AGR
'Mehmet had Can call his friend.'

3.3. DERIVATIONS

In an attempt to write lexical rules for all aspects of morphology, it is tempting to treat compounds in the same manner. Some compounds are predictable (10a–f). König (p.c.) suggests that these may be examples of prefixation, not compounding. This claim gains support for the fact that the modifier of the head is a function word in examples of this kind. It is possible to write lexical rules that refer to the syntactic information in the modifier (e.g., [cat=adposition]), but their semiproductivity must be accounted for by lexical semantic constraints on the head. Compounding of content words is much less semantically predictable. For instance, the modifier of the head *bez* in (11) is the instrument (11a), goal (11b), location (11c), theme (11d), or beneficiary (11e).

(10)

a.	*orta-boy*	b.	*orta-öğretim*
	middle-size		middle-education
	'medium'		'secondary education'
c.	*ön-yargı*	d.	*ön-söz*
	prior-judgment		prior-word
	'prejudice'		'preface'
e.	*dış-görünüş*	f.	*dış-güç*
	out-appearance		out-power
	'impression'		'foreign power'

(11)

a.	*ütü-bez-i*	b.	*el-bez-i*
	iron-cloth-COMP		hand-cloth-COMP
	'iron cloth'		'hand towel'
c.	*yer-bez-i*	d.	*toz-bez-i*
	floor-cloth-COMP		dust-cloth-COMP
	'mop'		'duster'
e.	*çocuk-bez-i*		
	child-cloth-COMP		
	'diaper'		

Compounding may also involve bracketing paradoxes which cannot be easily resolved by lexical rules. Example (12) (from (Göksel, 1993)) shows a case where the plural marker *-ler* semantically scopes over the nominal compound marker *-i*, but morphological bracketing is just the opposite. Göksel proposed two different specifications (surface and logical) for lexical items to account for this phenomenon. The paradox is resolved by combining *-ler* after the *logical* attachment of *-i* to the compound noun. The

difference in logical and surface attachment orders is not recoverable via lexical rules. We do not model compounds in this study.

(12) *otobüs bilet-ler-i*
bus ticket-PLU-COMP
'bus tickets'

Denominal verbs, deverbal nouns, and zero derivations are modelled respectively by adding subcategorization frames, discharging subcategorization frames, and type coercions, via lexical rules. The most difficult issue in derivations is semantic predictability. For instance, the *-CI* morpheme (with allomorphs *-cı/-ci/-cu/-cü/-çı/-çi/-çu/-çü*) adds the meaning "doer/user of something" (13a–b), "seller/lover of something" (13c), or habitual (13d).

(13)
a. *kamyon-cu*
truck-NtoN
'trucker, or freight operator'

b. *yol-cu*
road-NtoN
'traveller'

c. *şeker-ci*
candy-NtoN
'candy seller or lover'

d. *sabah-çı*
morning-NtoN
'morning person'

Clearly, this ambiguity cannot be resolved by lexical means without incorporating into lexical semantics a Qualia Structure (Pustejovsky, 1991), or lexical semantic constraints (Fass, 1993).[5] We have been incorporating these types of constraints. Unfortunately, descriptive work on Turkish linguistics in this regard is very scarce, and there is no ontology such as Levin's (1993) classification of English verbs. Using features like [∓animate], [∓artifact], [∓container], and [∓period], one can define semantic fields for the derivational morphemes. We expand the set of features as more lexical items are added to the lexicon. This is a very labor intensive task; the lack of a large scale initiative on Turkish lexicography in the manner of LDOCE or COBUILD is hindering the efforts for automatic extraction of lexical knowledge from on-line resources. Corpus-based studies on Turkish have only begun recently (e.g., (Oflazer and Tür, 1997)), and the largest tagged word list has around 22,000 entries, with raw texts around 2.5 million words. Clearly, these are not significant numbers to obtain a realistic frequency of attested word forms for an agglutinating language. Affix

[5]These two approaches assume that these notions can be lexicalized. As a counterpoint, Helmreich and Farwell (1999, this volume) argue that these constraints are pragmatic in nature; they cannot be handled by lexical sense extension rules. Incorporating detailed world knowledge into the lexicon most often compensates for the lack of a contextual model in an NLP system.

stripping and restoring the stems to lexical forms[6] have to be done before probabilistic techniques such as that of Briscoe and Copestake (1996) can be applied to tackle the semiproductivity of lexical rules for derivational morphology.

Our strategy is to obtain complex forms derivationally if the semantic relation of the bound morpheme to its stem is fairly predictable. We use lexicalized forms when the meaning is not compositional. This strategy provides a nonuniform treatment of derivational morphemes. For instance, we can derive nouns with *-CI* affix via a lexical rule when the stem is [+edible]. This restriction gives the predictable meaning 'one who likes x' where x is the meaning of the stem. This covers one sense of the derivation, but as (13c) shows, not all possible meanings can be derived by a lexical rule even though the stem satisfies the lexical semantic constraint. The other sense (one who sells x, for some edible thing x) is realized by a lexical item. Similarly, we use [+artifact] to control derivations by *-lIk*. For instance, although it is a derived noun, *göz-lük* (eye-NtoN "eyeglass") is considered a lexical item because no predictable property of the root warrants a lexical rule. But *göz-lük-lük* (eye-NtoN-NtoN) is derived by a lexical rule to obtain 'space for eyeglass (= eyeglass case)' because the stem is marked [+artifact] lexically. For many morphemes, this level of semiproductivity is not attainable. One such case is the denominal verb suffix *-le*, which is very productive but has no predictable meaning that can be derived from the lexical semantics of the stem. Zero derivations are modelled by lexical rules as well, e.g., the nominal use of adjectives is achieved by a single lexical item which may be interpreted as a term or a predicate by a lexical rule.

4. Evaluation of Lexical Rules

Butterworth (1983) proposed a Full Listing Hypothesis (FLH) in which all forms of a word's paradigm have lexical representation. He wrote:

> "I should point out a quite general problem with rules, namely what is to count as "regular" and hence rule governed. In English, as in other languages, there will be rules that apply to a large number of base forms, and rules that apply to smaller sets of base forms, and rules that apply to a singleton set (traditionally known as suppletive forms, e.g., *go*→ *went*; *be*→ *is*). So, in principle, any conjugation or declension CAN be generated by a rule, though rules which apply to only one LR [Lexical Representation] do not seem to confer much usefulness or generality

[6] An example of this is *kitap-ta* ("book-LOC"). When the inflectional suffix is stripped off, the effect of the morphophonemic process which maps the lexeme *kitab* to *kitap* by final stop devoicing must be undone.

> on this treatment. In any case, under a rule treatment, lexical entries would have to specify which rules are to apply in forming compounds [polymorphemic words], and it is possible to envisage a situation where the claimed advantages for this approach are lost by actually increasing the memory load to proportions comparable with the FLH model." (Butterworth, 1983, p.262)

In support of the full listing of even the rule governed patterns, he remarked:

> "It may be a tautology to claim that anything not available via a rule must be separately listed, it is by no means a tautology to claim that everything available via a rule must be available only in that way." (Butterworth, 1983, p.261):

As pointed out by Hankamer (1989), there is a recursive pattern in Turkish morphology for which a full-listing model would increase the memory load indefinitely. We believe that it can be accounted for only by a dynamic rule-based approach. Computationally, it clearly calls for run-time evaluation of the lexical rules. The phenomenon is known as *-ki* relativization.

The relative *-ki* suffix can be attached to nominal stems if the noun is in the genitive or in the locative case. (14a) and (14b) show the full paradigm, with corresponding examples in (14c) and (14d). Its semantics can be informally described as *'the one that x'* where x is the semantics of the case-marked nominal stem. The result is a nominal stem, and all noun inflections can be applied again. This leads to recursive expansions such as (14e) (this example may be referring to, say, a present in the box brought home by some friends). Examples like this are not uncommon among native speakers. Finding the closure of this relation during compilation leads to nontermination. Settling for a finite closure of *-ki* relativization would not be a principled solution to this phenomena.

(14) a. *meyve -ler -i -nde -ki*
fruit -PLU -POSS -LOC -REL

b. *arkadas -lar -ı -nın -ki*
friend -PLU -POSS -GEN -REL

c. *meyve-de-ki kurt*
fruit-LOC-REL worm
lit. 'the worm, the one in the fruit'

d. *Bu araba çalışmıyor ama arkadaş-ı-nın-ki çalışıyor*
this car not work but friend-POSS-GEN-REL works
lit. 'This car doesn't work but the car, the one that belongs to his/her friend, does.'

e. *ev* *-de* *-ki* *-ler* *-in* *-ki* *-nde* *-ki*
house -LOC -REL -PLU -GEN -REL -LOC -REL
lit. 'the one that is in the one that belongs to the ones that are in the house'

We have been testing our lexicon design as part of an HPSG grammar for Turkish written in ALE. Lexical rules utility of ALE has been used to implement inflectional morphology, and semantically predictable aspects of derivational morphology. The grammar development environment had to be modified to allow run-time evaluation of lexical rules. Compiling out the lexical rules off-line is quite impractical, since generating every possible form of roots in the lexicon causes exponential growth, or infinite expansion in the case of *-ki* relativization. For example, compilation of all surface forms for a lexicon of only 40 root forms produced around 2800 entries. Delayed evaluation (Johnson and Dorre, 1995; van Noord and Bouma, 1994), or precompiling the lexical rules for on-line expansion as suggested by Meurers and Minnen (1995), are viable options for avoiding off-line expansion. These techniques can either manually or automatically model rule interaction, and synthesize a proper ordering of rule execution. However, it is not clear to what extent the lexical rules *for morphology* can benefit from these techniques. There is not much rule interaction, and the constraints for evaluating the rules are instantiated by the time the stem is parsed and the affix is about to be attached. For these reasons, we allow run-time evaluation of the lexical rules without prior processing. Of course, run-time execution of rules puts the burden on parsing or generation. But this seems to be tolerable as far as morphological lexical rules are concerned.

5. Conclusion

For a language with rich morphology, morphological lexical rules can be used for grammatical and lexical meaning composition with varying degrees of productivity. Bracketing paradoxes and noncompositional nature of compounds render the lexical rule approach unsuitable for this aspect of morphology. Affixes with phrasal scope also put the lexical rules for morphology to the test. It is feasible to use a lexical rule for grammatical meaning composition in case the bracketing can be resolved by local information, such as the valence of the verbal stem. It remains to be seen whether all morphological markers for unbounded constructions (e.g., relativization, subordination) can be modelled by lexical rules that refer to the information in the stem. In either case, precompiling the Turkish lexicon is infeasible due to recursive morphological processes.

Regarding grammar architecture, the lexical rules approach to morphological operations situates morphology in a subcomponent of the lexicon,

perhaps not as an auxiliary as in the case of complete modularization, but nevertheless not as active a part of syntax as in the integrative approach. Although this idea is quite attractive, there is greater demand for lexical semantics to account for semiproductivity of derivational affixes. This is also true of parts of inflectional morphology in which lexicalized forms of some inflected words are attested.

Prominence of morphosyntax blurs the distinction between lexical and syntactic rules. Grammatical meaning composition encompasses both kinds of rules. The fact that some derivations can follow grammatical function changing affixes (e.g., *yap-tır-ım* do-CAUS-VtoN "sanction", and *at-ıl-gan* throw-PASS-VtoA "enterprising") indicates that at least *some* part of derivational morphology must be handled productively by lexical or syntactic rules.

The logical extreme of the lexical rule approach is handling all aspects of derivational morphology by lexical rules. But agglutinating nature of Turkish alone does not secure productive lexical meaning composition. Moreover, sense extension in derivations seems to depend on pragmatics as well as lexical conceptual structure (Helmreich and Farwell, 1999, this volume). Faced with the choice of shallow semantics with good but overgenerating morphological and syntactic coverage versus fine-grained semantics with constrained (attested) coverage, we chose the former due to lack of a comprehensive ontology. Another alternative was to adopt the full-listing approach for derivations only to obtain even finer-grained semantics, but given the extensive use of derivations in an agglutinating language, this did not seem plausible. Besides, in light of Helmreich and Farwell's critique of lexical rules, it seems more feasible to concentrate on building contextual models that can interact with (morpho)syntactic and semantic analyses, rather than trying to put everything in the lexicon.

Acknowledgements

Thanks to Ayşenur Birtürk, Aslı Göksel, Jorge Hankamer, Klaus von Heusinger, Wolf König, Stefan Müller, Kemal Oflazer, Meltem Turhan, and Ümit Turan for comments and discussion; and to Evelyne Viegas and two Kluwer reviewers for suggestions. All errors are, of course, our own. This research is supported in part by grants from NATO Science Division SfS III (grant no. TU-LANGUAGE), and Scientific and Technological Research Council of Turkey (grant no. EEEAG-90).

References

Anderson, S.R. (1992) *A-morphous Morphology.* Cambridge University Press.

Aronoff, M. (1976) *Word Formation in Generative Grammar*, Cambridge, MA: MIT Press.

Bozsahin, C. (1998) 'Deriving the Predicate-Argument Structure for a Free Word Order Language'. In: *Proceedings of the COLING-ACL'98*. Montreal, Canada.

Bozsahin, C. and E. Göçmen (1995) 'A Categorial Framework for Composition in Multiple Linguistic Domains'. In: *Proceedings of the Fourth International Conference on Cognitive Science of NLP (CSNLP'95)*. Dublin, Ireland.

Briscoe, T. and A. Copestake (1996) 'Controlling the Application of Lexical Rules'. In: E. Viegas (ed.): *Breadth and Depth of Semantic Lexicons*. SIGLEX96 Proceedings, Santa Cruz, California.

Butterworth, B. (1983) 'Lexical Representation'. In: *Language Production, vol. II*. London: Academic Press.

Carpenter, R. and G. Penn (1994) 'The Attribute Logic Engine User's Guide, Version 2.0'. Carnegie Mellon University, Pittsburgh.

Comrie, B. (1976) 'The syntax of causative constructions: cross-language similarities and divergences'. In: Shibatani and Masayoshi (eds.): *Syntax and Semantics 6: The grammar of causative constructions*. Academic Press.

Dowty, D. (1979) *Word Meaning and Montague Grammar*. Dordrecht: Kluwer.

Fass, D. (1993) 'Lexical Semantic Constraints'. In: J. Pustejovsky (ed.): *Semantics and the Lexicon*. Kluwer.

Göksel, A. (1993) 'Levels of Representation and Argument Structure in Turkish'. PhD Thesis, SOAS.

Güngördü, Z. and K. Oflazer (1995) 'Parsing Turkish Using the Lexical-Functional Grammar Formalism'. *Machine Translation* **10**, 293–319.

Hankamer, J. (1989) 'Morphological parsing and the Lexicon'. In: W. Marslen-Wilson (ed.): *Lexical Representation and Process*. MIT Press.

Helmreich, S. and D. Farwell (1999) '*Lexical Rules* is Italicized'. In: E. Viegas (ed.): *Breadth and Depth of Semantic Lexicons*. Kluwer Academic Press. (this volume)

Hoeksema, J. and R. D. Janda (1988) 'Implications of process-morphology for Categorial Grammar'. In: D. W. R.T. Oehrle, E. Bach (ed.): *Categorial Grammars and Natural Language Structures*. Dordrecht.

Johnson, M. and J. Dorre (1995) 'Memoization of Coroutined Constraints'. In: *Proceedings of the 33rd Annual Meeting of ACL*. Cambridge, MA.

Levin, B. (1993) *English verb classes and alternations: a preliminary investigation*. University of Chicago Press.

Lewis, G. L. (1967) *Turkish Grammar*. Oxford University Press.

Lieber, R. (1992) *Deconstructing Morphology*. University of Chicago Press.

Meurers, W. and G. Minnen (1995) 'A Computational Treatment of HPSG Lexical Rules as Covariation in Lexical Entries'. In: *Proceedings of the Fifth Int. Workshop on Natural Language Understanding and Logic Programming*. Lisbon.

van Noord, G. and G. Bouma (1994) 'Adjuncts and the Processing of Lexical Rules'. In: *Proceedings of COLING-94*. Kyoto.

Oflazer, K. and G. Tür (1997) 'Morphological Disambiguation by Voting Constraints'. In: *Proceedings of the 35th Annual Meeting of ACL*. Madrid, Spain.

Pollard, C. and I. A. Sag (1994) *Head-driven Phrase Structure Grammar*. CSLI Chicago.

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