

Morphology-Syntax Interface for Turkish LFG

Özlem Çetinöglü

Faculty of Engineering and Natural Sciences
Sabancı University
34956, Istanbul, Turkey
ozlemc@su.sabanciuniv.edu

Kemal Oflazer

Faculty of Engineering and Natural Sciences
Sabancı University
34956, Istanbul, Turkey
oflazer@sabanciuniv.edu

Abstract

This paper investigates the use of sublexical units as a solution to handling the complex morphology with productive derivational processes, in the development of a lexical functional grammar for Turkish. Such sublexical units make it possible to expose the internal structure of words with multiple derivations to the grammar rules in a uniform manner. This in turn leads to more succinct and manageable rules. Further, the semantics of the derivations can also be systematically reflected in a compositional way by constructing PRED values on the fly. We illustrate how we use sublexical units for handling simple productive derivational morphology and more interesting cases such as causativization, etc., which change verb valency. Our priority is to handle several linguistic phenomena in order to observe the effects of our approach on both the c-structure and the f-structure representation, and grammar writing, leaving the coverage and evaluation issues aside for the moment.

1 Introduction

This paper presents highlights of a large scale lexical functional grammar for Turkish that is being developed in the context of the ParGram project¹. In order to incorporate in a manageable way, the complex morphology and the syntactic relations mediated by morphological units, and to handle lexical representations of very productive derivations, we have opted to develop the grammar using sublexical units called *inflectional groups*.

Inflectional groups (IGs hereafter) represent the inflectional properties of segments of a complex

word structure separated by derivational boundaries. An IG is typically larger than a morpheme but smaller than a word (except when the word has no derivational morphology in which case the IG corresponds to the word). It turns out that it is the IGs that actually define syntactic relations between words. A grammar for Turkish that is based on words as units would have to refer to information encoded at arbitrary positions in words, making the task of the grammar writer much harder. On the other hand, treating morphemes as units in the grammar level implies that the grammar will have to know about morphotactics making either the morphological analyzer redundant, or repeating the information in the morphological analyzer at the grammar level which is not very desirable. IGs bring a certain form of normalization to the lexical representation of a language like Turkish, so that units in which the grammar rules refer to are simple enough to allow easy access to the information encoded in complex word structures.

That IGs delineate productive derivational processes in words necessitates a mechanism that reflects the effect of the derivations to semantic representations and valency changes. For instance, English LFG (Kaplan and Bresnan, 1982) represents derivations as a part of the lexicon; both *happy* and *happiness* are separately lexicalized. Lexicalized representations of adjectives such as *easy* and *easier* are related, so that both lexicalized and phrasal comparatives would have the same feature structure; *easier* would have the feature structure

- (1)
$$\left[\begin{array}{ll} \text{PRED} & \text{'easy'} \\ \text{ADJUNCT} & \left[\text{PRED} \text{ 'more'} \right] \\ \text{DEG-DIM} & \text{pos} \\ \text{DEGREE} & \text{comparative} \end{array} \right]$$

Encoding derivations in the lexicon could be applicable for languages with relatively unproductive derivational phenomena, but it certainly is not

¹<http://www2.parc.com/istl/groups/nlftt/pargram/>

possible to represent in the grammar lexicon,² all derived forms as lexemes for an agglutinative language like Turkish. Thus one needs to incorporate such derivational processes in a principled way along with the computation of the effects on derivations on the representation of the semantic information.

Lexical functional grammar (LFG) (Kaplan and Bresnan, 1982) is a theory representing the syntax in two parallel levels: Constituent structures (c-structures) have the form of context-free phrase structure trees. Functional structures (f-structures) are sets of pairs of attributes and values; attributes may be features, such as tense and gender, or functions, such as subject and object. C-structures define the syntactic representation and f-structures define more semantic representation. Therefore c-structures are more language specific whereas f-structures of the same phrase for different languages are expected to be similar to each other.

The remainder of the paper is organized as follows: Section 2 reviews the related work both on Turkish, and on issues similar to those addressed in this paper. Section 3 motivates and presents IGs while Section 4 explains how they are employed in a LFG setting. Section 5 summarizes the architecture and the current status of the our system. Finally we give conclusions in Section 6.

2 Related Work

Güngördü and Oflazer (1995) describes a rather extensive grammar for Turkish using the LFG formalism. Although this grammar had a good coverage and handled phenomena such as free-constituent order, the underlying implementation was based on pseudo-unification. But most crucially, it employed a rather standard approach to represent the lexical units: words with multiple nested derivations were represented with complex nested feature structures where linguistically relevant information could be embedded at unpredictable depths which made access to them in rules extremely complex and unwieldy.

Bozşahin (2002) employed morphemes overtly as lexical units in a CCG framework to account for a variety of linguistic phenomena in a prototype implementation. The drawback was that morphotactics was explicitly raised to the level of the sentence grammar, hence the categorial lexicon accounted for both constituent order and the morpheme order with no distinction. Oflazer's dependency parser (2003) used IGs as units between which dependency relations were established. Another parser based on IGs is Eryiğit and Oflazer's

²We use this term to distinguish the lexicon used by the morphological analyzer.

(2006) statistical dependency parser for Turkish. Çakıcı (2005), used relations between IG-based representations encoded within the Turkish Treebank (Oflazer et al., 2003) to automatically induce a CCG grammar lexicon for Turkish.

In a more general setting, Butt and King (2005) have handled the morphological causative in Urdu as a separate node in c-structure rules using LFG's restriction operator in semantic construction of causatives. Their approach is quite similar to ours yet differs in an important way: the rules explicitly use morphemes as constituents so it is not clear if this is just for this case, or all morphology is handled at the syntax level.

3 Inflectional Groups as Sublexical Units

Turkish is an agglutinative language where a sequence of inflectional and derivational morphemes get affixed to a root (Oflazer, 1994). At the syntax level, the unmarked constituent order is SOV, but constituent order may vary freely as demanded by the discourse context. Essentially all constituent orders are possible, especially at the main sentence level, with very minimal formal constraints. In written text however, the unmarked order is dominant at both the main sentence and embedded clause level.

Turkish morphotactics is quite complicated: a given word form may involve multiple derivations and the number of word forms one can generate from a nominal or verbal root is theoretically infinite. Turkish words found in typical text average about 3-4 morphemes including the stem, with an average of about 1.23 derivations per word, but given that certain noninflecting function words such as conjunctions, determiners, etc. are rather frequent, this number is rather close to 2 for inflecting word classes. Statistics from the Turkish Treebank indicate that for sentences ranging between 2 words to 40 words (with an average of about 8 words), the number of IGs range from 2 to 55 IGs (with an average of 10 IGs per sentence) (Eryiğit and Oflazer, 2006).

The morphological analysis of a word can be represented as a sequence of tags corresponding to the morphemes. In our morphological analyzer output, the tag \hat{DB} denotes derivation boundaries that we also use to define IGs. If we represent the morphological information in Turkish in the following general form:

$$\text{root} + \text{IG}_1 + \hat{DB} + \text{IG}_2 + \hat{DB} + \dots + \hat{DB} + \text{IG}_n.$$

then each IG_i denotes the relevant sequence of inflectional features including the part-of-speech for the root (in IG_1) and for any of the derived forms. A given word may have multiple such representations depending on any morphological ambiguity brought about by alternative segmentations of the

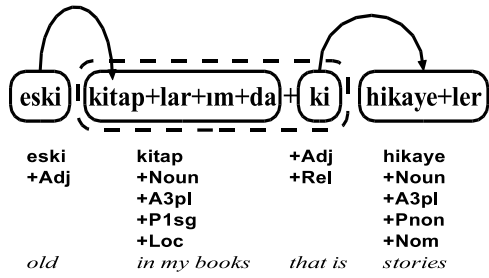


Figure 1: Modifier-head relations in the NP *eski kitaplarımdaki hikayeler*

word, and by ambiguous interpretations of morphemes.

For instance, the morphological analysis of the derived modifier *cezalandırılacak* (literally, “(the one) that will be given punishment”) would be:³

ceza (punishment) +Noun+A3sg+Pnon+Nom
 ^DB+Verb+Acquire
 ^DB+Verb+Caus
 ^DB+Verb+Pass+Pos
 ^DB+Adj+FutPart+Pnon

The five IGs in this word are:

1. +Noun+A3sg+Pnon+Nom
2. +Verb+Acquire
3. +Verb+Caus
4. +Verb+Pass+Pos
5. +Adj+FutPart+Pnon

The first IG indicates that the root is a singular noun with nominative case marker and no possessive marker. The second IG indicates a derivation into a verb whose semantics is “to acquire” the preceding noun. The third IG indicates that a causative verb (equivalent to “to punish” in English), is derived from the previous verb. The fourth IG indicates the derivation of a passive verb with positive polarity from the previous verb. Finally the last IG represents a derivation into future participle which will function as a modifier in the sentence.

The simple phrase *eski kitaplarımdaki hikayeler* (the stories in my old books) in Figure 1 will help clarify how IGs are involved in syntactic relations: Here, *eski* (old) modifies *kitap* (book) and not *hikayeler* (stories),⁴ and the locative phrase *eski*

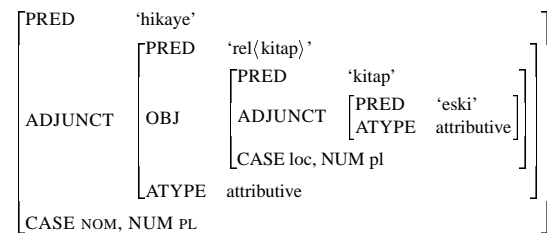
³The morphological features other than the obvious part-of-speech features are: +A3sg: 3sg number-person agreement, +Pnon: no possessive agreement, +Nom: Nominative case, +Acquire: acquire verb, +Caus: causative verb, +Pass: passive verb, +FutPart: Derived future participle, +Pos: Positive Polarity.

⁴Though looking at just the last POS of the words one sees an +Adj +Adj +Noun sequence which may imply that both adjectives modify the noun *hikayeler*

kitaplarımda (in my old books) modifies *hikayeler* with the help of derivational suffix *-ki*. Morpheme boundaries are represented by ‘+’ sign and morphemes in solid boxes actually define one IG. The dashed box around solid boxes is for word boundary. As the example indicates, IGs may consist of one or more morphemes.

Example (2) shows the corresponding f-structure for this NP. Supporting the dependency representation in Figure 1, f-structure of adjective *eski* is placed as the adjunct of *kitaplarımda*, at the innermost level. The semantics of the relative suffix *-ki* is shown as ‘rel(↑ OBJ)’ where the f-structure that represents the NP *eski kitaplarımda* is the OBJ of the derived adjective. The new f-structure with a PRED constructed on the fly, then modifies the noun *hikayeler*. The derived adjective behaves essentially like a lexical adjective. The effect of using IGs as the representative units can be explicitly seen in c-structure where each IG corresponds to a separate node as in Example (3).⁵ Here, DS stands for derivational suffix.

(2)



(3)

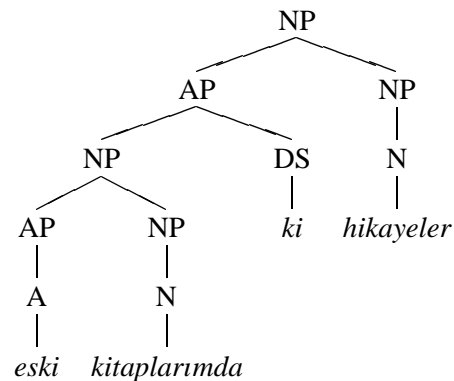


Figure 2 shows the modifier-head relations for a more complex example given in Example (4) where we observe a chain/hierarchy of relations between IGs

(4)

mavi renkli
 blue color-WITH
elbiselideki *kitap*
 dress-WITH-LOC-REL book

⁵Note that placing the sublexical units of a word in separate nodes goes against the Lexical Integrity principle of LFG (Dalrymple, 2001). The issue is currently being discussed within the LFG community (T. H. King, personal communication).

‘the book on the one with the blue colored dress’

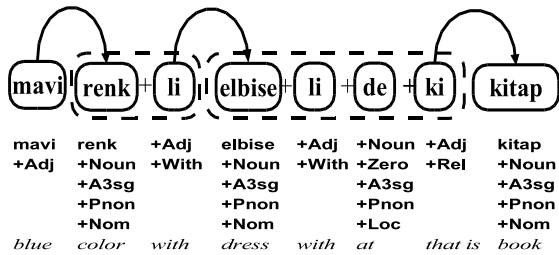


Figure 2: Syntactic Relations in the NP *mavi renkli elbiselideki kitap*

Examples (5) and (6) show respectively the constituent structure (c-structure) and the corresponding feature structure (f-structure) for this noun phrase. Within the tree representation, each IG corresponds to a separate node. Thus, the LFG grammar rules constructing the c-structures are coded using IGs as units of parsing. If an IG contains the root morpheme of a word, then the node corresponding to that IG is named as one of the syntactic category symbols. The rest of the IGs are given the node name DS (to indicate derivational suffix), no matter what the content of the IG is.

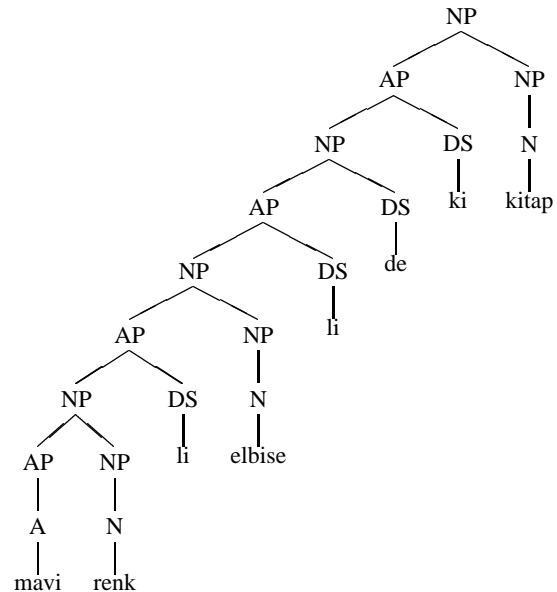
The semantic representation of derivational suffixes plays an important role in f-structure construction. In almost all cases, each derivation that is induced by an overt or a covert affix gets a OBJ feature which is then unified with the f-structure of the preceding stem already constructed, to obtain the feature structure of the derived form, with the PRED of the derived form being constructed on the fly. A PRED feature thus constructed however is not meant to necessarily have a precise lexical semantics. Most derivational suffixes have a consistent (lexical) semantics⁶, but some don't, that is, the precise additional lexical semantics that the derivational suffix brings in, depends on the stem it is affixed to. Nevertheless, we represent both cases in the same manner, leaving the determination of the precise lexical semantics aside.

If we consider Figure 2 in terms of dependency relations, the adjective *mavi* (blue) modifies the noun *renk* (color) and then the derivational suffix *-li* (with) kicks in although the *-li* is attached to *renk* only. Therefore, the semantics of the phrase should be *with(blue color)*, not *blue with(color)*. With the approach we take, this difference can easily be represented in both the f-structure as in the leftmost branch in Example (5)

⁶e.g., the “to acquire” example earlier

and the c-structure as in the middle ADJUNCT f-structure in Example (6). Each DS in c-structure gives rise to an OBJect in c-structure. More precisely, a derived phrase is always represented as a binary tree where the right daughter is always a DS. In f-structure DS unifies with the mother f-structure and inserts PRED feature which subcategorizes for a OBJ. The left daughter of the binary tree is the original form of the phrase that is derived, and it unifies with the OBJ of the mother f-structure.

(5)



4 Inflectional Groups in Practice

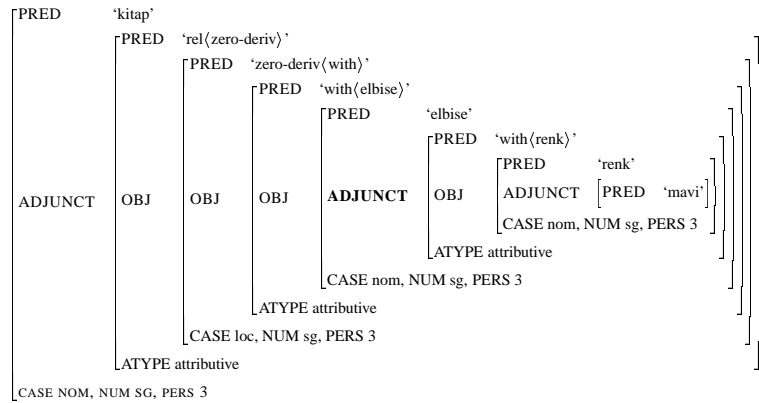
We have already seen how the IGs are used to construct on the fly PRED features that reflect the lexical semantics of the derivation. In this section we describe how we handle phenomena where the derivational suffix in question does not explicitly affect the semantic representation in PRED feature but determines the semantic role so as to unify the derived form or its components with the appropriate external f-structure.

4.1 Sentential Complements and Adjuncts, and Relative Clauses

In Turkish, sentential complements and adjuncts are marked by productive verbal derivations into nominals (infinitives, participles) or adverbials, while relative clauses with subject and non-subject (object or adjunct) gaps are formed by participles which function as adjectivals modifying a head noun.

Example (7) shows a simple sentence that will be used in the following examples.

(6)

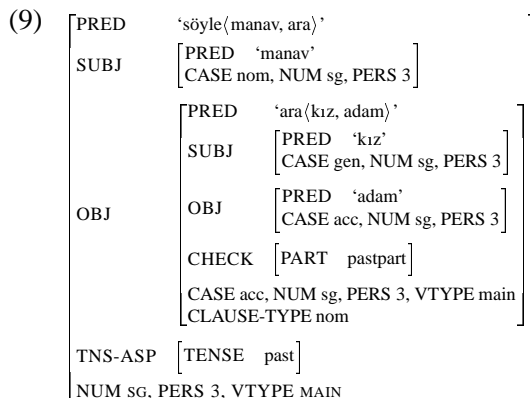


(7) *Kız adamı aradı.*
 Girl-NOM man-ACC call-PAST
 'The girl called the man'

In (8), we see a past-participle form heading a sentential complement functioning as an object for the verb *söyledi* (said).

(8) *Manav kızın adamı aradığını söyledi.*
 Grocer-NOM girl-GEN man-ACC
 call-PASTPART-ACC say-PAST
 'The grocer said that the girl called the man'

Once the grammar encounters such a sentential complement, everything up to the participle IG is parsed, as a normal sentence and then the participle IG appends nominal features, e.g., CASE, to the existing f-structure. The final f-structure is for a noun phrase, which now is the object of the matrix verb, as shown in Example (9). Since the participle IG has the right set of syntactic features of a noun, no new rules are needed to incorporate the derived f-structure to the rest of the grammar, that is, the derived phrase can be used as if it is a simple NP within the rules. The same mechanism is used for all kinds of verbal derivations into infinitives, adverbial adjuncts, including those derivations encoded by lexical reduplications identified by multi-word construct processors.



Relative clauses also admit to a similar mechanism. Relative clauses in Turkish are gapped sentences which function as modifiers of nominal heads. Turkish relative clauses have been previously studied (Barker et al., 1990; Güngördü and Engdahl, 1998) and found to pose interesting issues for linguistic and computational modeling. Our aim here is not to address this problem in its generality but show with a simple example, how our treatment of IGs encoding derived forms handle the mechanics of generating f-structures for such cases.

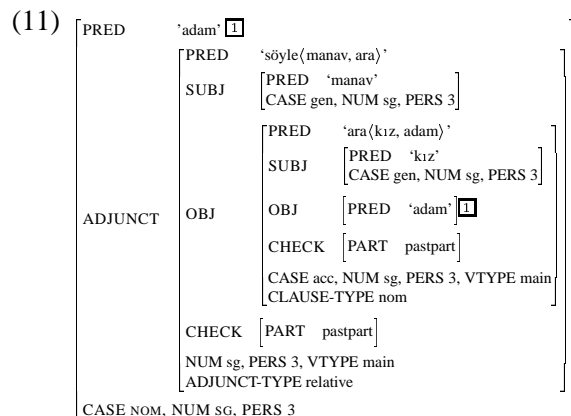
Kaplan and Zaenen (1988) have suggested a general approach for handling long distance dependencies. They have extended the LFG notation and allowed regular expressions in place of simple attributes within f-structure constraints so that phenomena requiring infinite disjunctive enumeration can be described with a finite expression. We basically follow this approach and once we derive the participle phrase we unify it with the appropriate argument of the verb using rules based on functional uncertainty. Example (10) shows a relative clause where a participle form is used as a modifier of a head noun, *adam* in this case.

(10) *Manavın kızın []_i*
 Grocer-GEN girl-GEN obj-gap
aradığını söylediği adamı
 call-PASTPART-ACC say-PASTPART man-NOM
 'The man the grocer said the girl called'

This time, the sentence is parsed with a gap with an appropriate functional uncertainty constraint, and when the participle IG is encountered the sentence f-structure is derived into an adjective and the gap in the derived form, the object here, is then unified with the head word as marked with co-indexation in Example (11).

The example sentence (10) includes Example (8) as a relative clause with the object extracted, hence the similarity in the f-structures can be observed easily. The ADJUNCT in Example (11)

is almost the same as the whole f-structure of Example (9), differing in TNS-ASP and ADJUNCT-TYPE features. At the grammar level, both the relative clause and the complete sentence is parsed with the same core sentence rule. To understand whether the core sentence is a complete sentence or not, the finite verb requirement is checked. Since the requirement is met by the existence of TENSE feature, Example (8) is parsed as a complete sentence. Indeed the relative clause also includes temporal information as ‘pastpart’ value of PART feature, of the ADJUNCT f-structure, denoting a past event.



4.2 Causatives

Turkish verbal morphotactics allows the production multiply causative forms for verbs.⁷ Such verb formations are also treated as verbal derivations and hence define IGs. For instance, the morphological analysis for the verb *aradı* (s/he called) is

ara+Verb+Pos+Past+A3sg

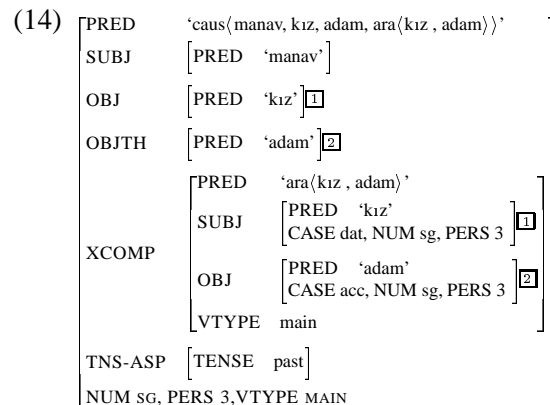
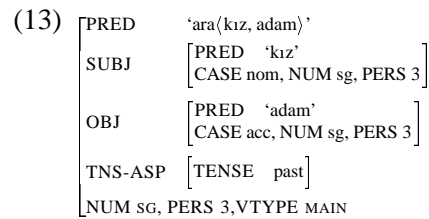
and for its causative *arattı* (s/he made (someone else) call) the analysis is

ara+Verb^{DB}+Verb+Caus+Pos+Past+A3sg.

In Example (12) we see a sentence and its causative form followed by respective f-structures for these sentences in Examples (13) and (14). The detailed morphological analyses of the verbs are given to emphasize the morphosyntactic relation between the bare and causativized versions of the verb.

- (12)
- a. *Kız adamı aradı.*
 Girl-NOM man-ACC call-PAST
 ‘The girl called the man’
- b. *Manav kıza adamı arattı.*
 Grocer-NOM girl-DAT man-ACC
 call-CAUS-PAST
 ‘The grocer made the girl call the man’

⁷Passive, reflexive, reciprocal/collective verb formations are also handled in morphology, though the latter two are not productive due to semantic constraints. On the other hand it is possible for a verb to have multiple causative markers, though in practice 2-3 seem to be the maximum observed.



The end-result of processing an IG which has a verb with a causative form is to create a larger f-structure whose PRED feature has a SUBJECT, an OBJECT and a XCOMPLEMENT. The f-structure of the first verb is the complement in the f-structure of the causative form, that is, its whole structure is embedded into the mother f-structure in an encapsulated way. The object of the causative (causee - that who is caused by the causer – the subject of the causative verb) is unified with the subject of the inner f-structure. If the original verb is transitive, the object of the original verb is further unified with the OBJTH of the causative verb. All of grammatical functions in the inner f-structure, namely XCOMP, are also represented in the mother f-structure and are placed as arguments of caus since the flat representation is required to enable free word order in sentence level.

Though not explicit in the sample f-structures, the important part is unifying the object and former subject with appropriate case markers, since the functions of the phrases in the sentence are decided with the help of case markers due to free word order. If the verb that is causativized subcategorizes for an direct object in accusative case, after causative formation, the new object unified with the subject of the causativized verb should be in dative case (Example 15). But if the verb in question subcategorizes for a dative or an ablative oblique object, then this object will be transformed into a direct object in accusative case after causativization (Example 16). That is, the causativization will select the case of the object of the causative verb, so as not to “interfere” with the object of the verb that is causativized. In causativized intransitive verbs the causative object is always in accusative case.

- (15) a. *adam kadını aradı.*
man-NOM woman-ACC call-PAST
'the man called the woman'
- b. *adama kadını arattı.*
man-DAT woman-ACC call-CAUS-PAST
'(s/he) made the man call the woman'
- (16) a. *adam kadına vurdu.*
man-NOM woman-DAT hit-PAST
'the man hit the woman'
- b. *adama kadına vurdurdu.*
man-ACC woman-DAT hit-CAUS-PAST
'(s/he) made the man hit the woman'

All other derivational phenomena can be solved in a similar way by establishing the appropriate semantic representation for the derived IG and its effect on the semantic representation.

5 Current Implementation

The implementation of the Turkish LFG grammar is based on the Xerox Linguistic Environment (XLE) (Maxwell III and Kaplan, 1996), a grammar development platform that facilitates the integration of various modules, such as tokenizers, finite-state morphological analyzers, and lexicons. We have integrated into XLE, a series of finite state transducers for morphological analysis and for multi-word processing for handling lexicalized, semi-lexicalized collocations and a limited form of non-lexicalized collocations.

The finite state modules provide the relevant ambiguous morphological interpretations for words and their split into IGs, but do not provide syntactically relevant semantic and subcategorization information for root words. Such information is encoded in a lexicon of root words on the grammar side.

The grammar developed so far addresses many important aspects ranging from free constituent order, subject and non-subject extractions, all kinds of subordinate clauses mediated by derivational morphology and has a very wide coverage NP subgrammar. As we have also emphasized earlier, the actual grammar rules are oblivious to the source of the IGs, so that the same rule handles an adjective - noun phrase regardless of whether the adjective is lexical or a derived one. So all such relations in Figure 2⁸ are handled with the same phrase structure rule.

The grammar is however lacking the treatment of certain interesting features of Turkish such as *suspended affixation* (Kabak, 2007) in which the inflectional features of the last element in a coordination have a phrasal scope, that is, all other

⁸Except the last one which requires some additional treatment with respect to definiteness.

coordinated constituents have certain default features which are then "overridden" by the features of the last element in the coordination. A very simple case of such suspended affixation is exemplified in (17a) and (17b). Note that although this is not due to derivational morphology that we have emphasized in the previous examples, it is due to a more general nature of morphology in which affixes may have phrasal scopes.

- (17) a. *kız adam ve kadını aradı.*
girl man-NOM and woman-ACC
call-PAST
'the girl called the man and the woman'
- b. *kız [adam ve kadın]-ı aradı.*
girl [man and woman]-ACC call-PAST
'the girl called the man and the woman'

Suspended affixation is an example of a phenomenon that IGs do not seem directly suitable for. The unification of the coordinated IGs have to be done in a way in which non-default features of the final constituent is percolated to the upper node in the tree as is usually done with phrase structure grammars but unlike coordination is handled in such grammars.

6 Conclusions and Future Work

This paper has described the highlights of our work on developing a LFG grammar for Turkish employing sublexical constituents, that we have called inflectional groups. Such a sublexical constituent choice has enabled us to handle the very productive derivational morphology in Turkish in a rather principled way and has made the grammar more or less oblivious to morphological complexity.

Our current and future work involves extending the coverage of the grammar and lexicon as we have so far included in the grammar lexicon only a small subset of the root lexicon of the morphological analyzer, annotated with the semantic and subcategorization features relevant to the linguistic phenomena that we have handled. We also intend to use the Turkish Treebank (Oflazer et al., 2003), as a resource to extract statistical information along the lines of Frank et al. (2003) and O'Donovan et al. (2005).

Acknowledgement

This work is supported by TUBITAK (The Scientific and Technical Research Council of Turkey) by grant 105E021.

References

- Chris Barker, Jorge Hankamer, and John Moore, 1990. *Grammatical Relations*, chapter Wa and Ga in Turkish. CSLI.
- Cem Bozşahin. 2002. The combinatory morphemic lexicon. *Computational Linguistics*, 28(2):145–186.
- Miriam Butt and Tracey Holloway King. 2005. Restriction for morphological valency alternations: The Urdu causative. In *Proceedings of The 10th International LFG Conference*, Bergen, Norway. CSLI Publications.
- Ruken Çakıcı. 2005. Automatic induction of a CCG grammar for Turkish. In *Proceedings of the ACL Student Research Workshop*, pages 73–78, Ann Arbor, Michigan, June. Association for Computational Linguistics.
- Mary Dalrymple. 2001. *Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*. Academic Press, New York.
- Gülşen Eryiğit and Kemal Oflazer. 2006. Statistical dependency parsing for turkish. In *Proceedings of EACL 2006 - The 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy. Association for Computational Linguistics.
- Anette Frank, Louisa Sadler, Josef van Genabith, and Andy Way. 2003. From treebank resources to LFG f-structures:automatic f-structure annotation of treebank trees and CFGs extracted from treebanks. In Anne Abeille, editor, *Treebanks*. Kluwer Academic Publishers, Dordrecht.
- Zelal Güngördü and Elisabeth Engdahl. 1998. A relational approach to relativization in Turkish. In *Joint Conference on Formal Grammar, HPSG and Categorical Grammar*, Saarbrücken, Germany, August.
- Zelal Güngördü and Kemal Oflazer. 1995. Parsing Turkish using the Lexical Functional Grammar formalism. *Machine Translation*, 10(4):515–544.
- Bariş Kabak. 2007. Turkish suspended affixation. *Linguistics*, 45. (to appear).
- Ronald M. Kaplan and Joan Bresnan. 1982. Lexical-functional grammar: A formal system for grammatical representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, pages 173–281. MIT Press, Cambridge, MA.
- Ronald M. Kaplan and Annie Zaenen. 1988. Long-distance dependencies, constituent structure, and functional uncertainty. In M. Baitin and A. Kroch, editors, *Alternative Conceptions of Phrase Structure*. University of Chicago Press, Chicago.
- John T. Maxwell III and Ronald M. Kaplan. 1996. An efficient parser for LFG. In Miriam Butt and Tracy Holloway King, editors, *The Proceedings of the LFG '96 Conference*, Rank Xerox, Grenoble.
- Ruth O'Donovan, Michael Burke, Aoife Cahill, Josef van Genabith, and Andy Way. 2005. Large-scale induction and evaluation of lexical resources from the Penn-II and Penn-III Treebanks. *Computational Linguistics*, 31(3):329–365.
- Kemal Oflazer, Bilge Say, Dilek Zeynep Hakkani-Tür, and Gökhan Tür. 2003. Building a Turkish treebank. In Anne Abeille, editor, *Building and Exploiting Syntactically-annotated Corpora*. Kluwer Academic Publishers.
- Kemal Oflazer. 1994. Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9(2):137–148.
- Kemal Oflazer. 2003. Dependency parsing with an extended finite-state approach. *Computational Linguistics*, 29(4):515–544.