Two–level Description of Turkish Morphology

Kemal Oflazer
Department of Computer Engineering and Information Science
Bilkent University, Ankara, Turkey
Fax: (90-4) 266 4127 e-mail: ko@trbilun.bitnet

1 Introduction

This poster paper describes a full scale two-level morphological description (Karttunen, 1983, Koskenniemi, 1983) of Turkish word structures. The description has been implemented using the PC-KIMMO environment (Antworth, 1990) and is based on a root word lexicon of about 23,000 roots words. Almost all the special cases of and exceptions to phonological and morphological rules have been implemented.

Turkish is an agglutinative language with word structures formed by productive affixations of derivational and inflectional suffixes to root words. Turkish has finite-state but nevertheless rather complex morphotactics. Morphemes added to a root word or a stem can convert the word from a nominal to a verbal structure or vice-versa, or can create adverbial constructs. The surface realizations of morphological constructions are constrained and modified by a number of phonetic rules such as vowel harmony.

2 Two-level description of Turkish morphology

The phonetic rules of contemporary Turkish have been encoded using 22 two-level rules while the morphotactics of the agglutinative word structures has been encoded as finite-state machines for verbal, nominal paradigms. Our lexicons are based on the comprehensive word list that we have compiled for our spelling checker developed earlier (Solak and Oflazer, 1992). We have lexicons for *nouns, adjectives verbs, compound nouns, proper nouns, pronouns, adverbs, connectives, exclamations, postpositions, acronyms, technical words, special cases,* There are total of 18,500 nominal (nouns + adjectives) roots and about 2,450 verbal roots. There are about 100 lexicons for suffixes.

3 Example Output

Here we provide a sample output from our implementation (slightly edited for proper orthography):

Input Morpheme Struct.	Gloss *English meaning*
çalışmanın	
çalış+mA+Hn +nHn	V(çalış)+VtoN(ma)+2PS-POS+GEN *of your work(ing)*
çalış+mA+nHn	V(çalış)+VtoN(ma)+GEN *of the work(ing)*
çocuğu	
çocuk+sH	N(çocuk)+3PS-POS *his/her child*
çocuk+yH	N(çocuk)+ACC *child (accusative)*
alınmış	
al+Hn+ymHş	N(al)+2PS-POS+NtoV()+NARR+3PS *(it) was your red (one)*
al+nHn+ymHş	N(al)+GEN+NtoV()+NARR+3PS *(it) belongs to the red (one)*
al$ın+ymHş	N(alın)+NtoV()+NARR+3PS *(it) was a forehead*
al+Hn+mHş	V(al)+PASS+VtoAdj(mis) *(a) taken (object)*
al+Hn+mHş	V(al)+PASS+NARR+3PS *it was taken*
alın+mHş	V(alın)+VtoAdj(mis) *(an) offended (person)*
alın+mHş	V(alın)+NARR+3PS *s/he was offended*
boynu	
boy$un+sH	N(boyun)+3PS-POS *(his/her) neck*
boy$un+yH	N(boyun)+ACC *neck (accusative)*

4 Conclusions

This poster has presented a summary of the first full scale implementation of two-level description of Turkish morphology. We have been using this description as a morphological parsing module in a number of applications like LFG parsing, ATN parsing and semantics analysis of Turkish sentences.

References

[Antworth, 1990] Evan L. Antworth. *PC-KIMMO: A two-level processor for Morphological Analysis.* Summer Institute of Linguistics, Dallas, Texas, 1990.

[Karttunen, 1983] Lauri Karttunen. KIMMO: A general morphological processor. *Texas Linguistic Forum,* 22:163 - 186, 1983.

[Koskenniemi, 1983] Kimmo Koskenniemi. Two-level morphology: A general computational model for word form recognition and production. Publication No: 11, Department of General Linguistics, University of Helsin, 1983.

[Solak and Oflazer, 1992] Ayşın Solak and Kemal Oflazer. Parsing agglutinative word structures and its application to spelling checking for Turkish. In *Proceedings of the* 15^{th} *International Conference on Computational Linguistics,* volume 1, pages 39 - 45, Nantes, France, 1992. International Commitee on Computational Linguistics.