Selection of Linker Type in Emphatic Reduplication: Speaker's Intuition meets Corpus Statistics

Özkan Kılıç (ozkan.kilic@gmail.com)

Cem Bozşahin (bozsahin@metu.edu.tr)
Department of Cognitive Science, Graduate School of Informatics
Middle East Technical University (METU), Ankara, Turkey

Abstract

Turkish Emphatic Reduplication (TER) occurs in adjectives and adverbs to accentuate their meanings. The current experimental study to investigate the selection of the linker type in TER indicated that responses from the participants correlate with some lexical statistics. The result relies on the statistics from a corpus with approximately 2 million Turkish words, which we use in lieu of lexical statistics. The frequency order of linker choice reported by the participants was exactly the opposite of the order of frequency of words in which the linker and the first consonant co-occur consecutively in the corpus. Such a direct link to lexicon was unexpected. We suggest that TER, an apparently phonological operation, depends on lexical access for selecting the appropriate linker whose co-occurrence with the initial consonant of the reduplicated word is infrequent. Our results relate morphology and lexicon in more ways than the blocking phenomena, and suggest that TER may be morpholexical.

Keywords: Morphology; emphatic reduplication; lexicon; lexical frequency.

Introduction

Turkish is generally considered to be a language with subject-object-verb word order, and it is morphologically agglutinating, with considerably involved morphology. The Turkish word structure depends heavily on suffixes, which are transparently stacked on the leftmost stem like "beads on a string" (see Kornfilt, 1997; Lewis, 2000; Göksel & Kerslake, 2005 for a detailed review of Turkish). However, not all Turkish morphology is suffixal, and it is not always the case that we can enumerate the allomorphies clearly in non-suffixal morphology and phonology, with some good variation among speakers, which is not the case for suffixal morphology. Perhaps not surprisingly, these cases involve items which might appear to look like prefixes. We suggest a way to relate these differences in the processing of bound items to lexical statistics, as evidence for another case of interaction between morphology and lexicon.

Turkish Emphatic Reduplication, henceforth TER, is one such resource showing greater speaker variation. It is a derivational process which intensifies the meaning of adjectives and some adverbs. Phonologically it involves the duplication of the initial *(C)V* of the base, and addition of a prefix-like item as a linker to the root, which is a consonant from the set {*p, s, m, r*}. (Demircan, 1987; Oztaner 1996; Wedel, 1999; Yu, 1999; Kelepir, 2000; Kim, 2007; Dhillon, 2009). All words beginning with a vowel are infixed with *-p-* as the linker. In some cases the *(C)V+linker* "prefix" is also followed by an additional "infix" from the set {*-A, -Il, -Am*} as in some of the examples below.

(1) ka-s-katı — çı-r-ıl-çıplak
RED solid — RED naked
'hard as a rock' — 'totally naked'

pa-r-am-parça — dü-p-e-düz
RED torn — RED plain
'completely torn apart' — 'utterly'

Related Work

Inkelas & Zoll (2005) use cophonology to explain emphatic reduplication. A cophonology is a morphological function associated with particular morphological constructions to model morphologically conditioned phonology. Cophonologies receive words or morphemes as input, and perform some operations such as constraint ranking, truncation, and velar deletion on the input to be sent to the phonological interface (Inkelas & Orgun, 1995; Inkelas, 2000). Truncation and addition act on the word *beyaz* 'white' to produce *bem−*. Then, the mother node links the subconstituent daughters to the input and shifts stress to the truncated one to form *bémbeyaz* 'snow white'. Demircan (1987), and later Wedel (1999, 2000), examined TER as a phonological operation and summarized the linker selection constraints as follows:

(2) (i) The linker from the set {*p, s, m, r*} cannot be identical with the initial consonant (C_1) of the base: *pembe* 'pink' → **peppembe*, although $p \in$ {*p, s, m, r*}. *Perpembe* is possible but not likely (see below).

(ii) The linker cannot be identical to the second consonant (C_2) of the base: *pembe* → **pempembe*/*pespembe*, although $m \in$ {*p, s, m, r*}.

(iii) The phonetic features {coronal, sonorant, labial, continuant} of the linker cannot be identical with those of the second segment of the base. The linker with the most contrasting features is selected for perceptual salience.

(iv) The linker is selected in a way that it can establish an optimization or balance among the features contributing to the featural contrast with respect to base.

Some examples in Turkish seem to be orthogonal to these constraints. For example, in addition to *çı-r-ıl-çıplak*, which is the commonly assumed reduplicated form of *çıplak* as in (1), *çı-s-çıplak*, *çı-r-çıplak*, *çı-m-çıplak* and *çı-p-çıplak* do occur in the Web.

The constraints in (2i–iv) can be violated when phonological productivity is put to the test in relation to morphology and the lexicon. We asked 50 participants to reduplicate *pırasa* 'leek', and along with the expected *pımpırasa* and *pıspırasa*, we also received *pırpırasa* and *pıppırasa*, in more or less equal distribution.

In this paper we report our experimental study which investigates the selection of the linker type, and its relation to morphology and the lexicon. The results indicate that responses from participants show disagreements with the literature about familiar words which are known to be non-TER targets. It also suggests that TER may be morpholexical, which is quite contrary to current literature.

The database for our statistics is a corpus of approximately 2 million Turkish words. We consider the database to be a rough approximate of a native speaker's lexical statistics. We asked the participants to emphatically reduplicate some words from this corpus. We asked them to play a "what-if" game to reduplicate the words that they would normally resist to reduplicate, such as *masa* 'table'. Among 50 participants and 31 target words, only one target, the only one that begins with a vowel, showed unanimous agreement, all others showing varying degrees of agreement. We present our method and the experiments toward understanding these results.

TER vs. Reduplication

Kim (2007) points out that the most productive "prefixation" in Turkish is observed in reduplication. According to Göksel & Kerslake (2005), Turkish duplication can be observed in three ways: M-reduplication, doubling, and TER, as explained above. We show below that, unlike TER, the first two kind are phonological-syntactic operations, therefore not operating on morphological properties but phonological-syntactic ones. In this regard they are expected to be less susceptible to lexical statistics.

M-reduplication

If a word or compound to be m-reduplicated starts with a vowel, the original word is prefixed with *m-*, and then duplicated as shown in (3a). If it starts with a consonant other than *m-*, the consonant is replaced with *m-*, and the new form is duplicated as shown in (3b). In case the word or the compound starts with *m-*, it is followed by the word *falan* 'like, so and so'. M-reduplications can occur in all syntactic positions.

(3) a. [Çocuklar]$_{NP}$ [[akıcı makıcı]$_{ADV}$ [konuşmazlar]$_{V}$]$_{VP}$
child-PLU fluent M-DUP speak-NEG-AOR-3PL
lit. 'Children do not speak fluently (and the like)'

b. [Çocuklar mocuklar]$_{NP}$ [[akıcı]$_{ADV}$ [konuşmazlar]$_{V}$]$_{VP}$
child-PLU M-DUP fluent speak-NEG-AOR-3PL
lit. 'Children (and the like) do not speak fluently'

We note that the results of this process are two independent words, both phonologically and syntactically. We can, for example, choose the duplicated form in (3b) as the target of a construction:

(4) a. Çocuklar mocuklar akıcı konuşmazlar.
childPLU M-DUP fluently speak-NEG-AOR-3PL
'Children (and the like) do not speak fluently'

b. Mocuklar hiç konuşmazlar.
M-DUP never speak-NEG-AOR-3PL
lit. 'The likes do not speak at all.'

Doubling

Doubling occurs in two ways: simple doubling, and doubling in lexical formations. In simple doubling, the word is repeated, as in (5). Depending on the syntactic category of the targeted lexeme, it can produce adverbials, adjectivals and measure terms (Göksel & Kerslake 2005).

(5) tek tek zaman zaman
one DUP time DUP
'one by one' 'time to time'

Some additional morphemes, such as the plural suffix and the question particle (QP), are attached to the sister constituents as in (6a) and (6b), or one of the constituents undergoes phonetic changes, as in (6c), for doubling in lexical formations.

(6) a. güzel-ler güzel-i bir kız
beautiful-PLU beautiful-POSS a girl
'a very beautiful girl'

b. güzel mi güzel bir kız
beautiful QP beautiful a girl
'a very beautiful girl'

c. ufak tefek bir kutu
little ϕ_i(little) a box
'a tiny box'

Among these alternatives, the last one seems closest to a morphological-lexical operation. ϕ_i in (6c) stands for cophonology (Orgun, 1996; Orgun, 1999; Inkelas & Zoll, 2005), which is the morphological function associated with particular morphological constructions to model morphologically conditioned phonology. The basic idea is shown in Figure 1.

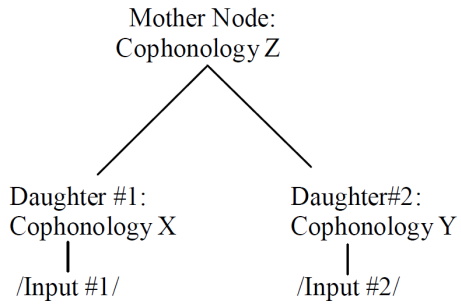

Figure 1: Template for cophonologies.

Inkelas & Zoll (2005) employ cophonologies in their Morphological Doubling Theory (MDT), and stress that the theory is morphologically motivated because it makes use of roots, morphs and affixes, rather than mora, coda or foot. The model works in a binary manner, in which there are two inputs called daughter nodes, and the output in the tree's root is called the mother node. In MDT, the reduplicant and base are both generated by morphology as part of a construction that also embodies semantic and phonological generalizations concerning the output of reduplication (Inkelas, 2005).

However, Göksel & Kerslake (2005:101) classify lexical formations as idiomatic expressions, and our intuitions are consistent with this observation. Take for example the following from the same page in their book:

(7) a. konu komşu
LEX-FORM neighbor
'neighbors', lit. 'neighbor-neighbor'

b. süklüm püklüm
crestfallen LEX-FORM
'in a crestfallen manner'

Firstly, they are not always right-headed or left-headed, which seems a bit unusual for a purportedly morphological operation. Secondly, as Göksel & Kerslake (2005) point out, the copy may or may not exist independently. We add to this an additional observation that the doubled word is indeed a phonological word, not a suffix, prefix, and not necessarily a lexeme.

It is not clear to us whether examples such as (6c) should be made part of an experiment on lexical statistics versus speaker production of *bound* elements, because it is not clear whether something other than multi-word word formation is involved here, although we are quite certain that independent words should be left out, i.e. examples such as (6a–b). We chose to take idiomatic lexical formations (6c) outside the scope of our experiments too, and concentrate only on TER.

In Morphological Doubling Theory, MDT, the features given above can be ranked by the cophonologies to determine the linker in the emphatically reduplicated form of TER. Yu (1999) argued that the allomorphy in Turkish reduplication could be accounted for by positing morphotactic constraints, which spell out the form of each of the allomorphs that dominate certain phonotactic constraints. The ultimate selection of the appropriate morph depends on the harmonic satisfaction of the lower-ranked phonotactic constraints on the linker.

Demircan (1987) analyzed 121 emphatically reduplicated adjectives, and concluded that the frequency of reduplicated adjectives showed the ranking *-p* > *-m* > *-s* > *-r*, in which *-p* is the most selected linker, and *-r* the least. Of all the adjectives, 46% are reduplicated with *-p*, 29% with *-m*, 18% with *-s*, and 7% with *-r*. In another study, Wedel (1999, 2000) concluded that TER with the linker *-r* might be lexicalized.

In contrast with these findings, and with the constraints summarized in (2), we noted some frequently occurring "exceptions" in §1, for example, among others, *çı-r-ıl-çıplak*, which is the expected reduplicated form of *çıplak*, but also *çı-s-çıplak*, *çı-m-çıplak*, *çı-r-çıplak* and *çı-p-çıplak*.

There seems to be no easy generalization across the speakers about how they emphatically reduplicate a novel word. Auspiciously, there are tendencies depending on the linker type, with respect to lexical co-occurrence frequencies, which we explain below.

Method and Findings

In order to thoroughly understand the linker selection choices available to native speakers of Turkish, a questionnaire consisting of 31 nonadjectival words composed of Turkish nouns and verbs was prepared. The word list was given to 25 male and 25 female participants, all university graduates, average age 34.20, and all native speakers of Turkish. Nonadjectives were deliberately selected to guarantee that the participants would be very unlikely to have applied TER to the words in the list before. The participants were asked the following questions: if the words were to be treated as adjectives, how would they emphatically reduplicate them? They were allowed to give single word answers. We also asked whether they had ever reduplicated the words, how they knew how to reduplicate the words, and the average time in seconds it took them to reduplicate each word. Because our plans did not involve reaction time experiments, we considered the last piece of information as the perceived difficulty and/or effort on behalf of the speaker. About two-thirds of the subjects took the experiment in our presence, and our timing seems to concur what they reported.

We have been told that they had never reduplicated any of the words before. For their own explanation of self performance, the participants responded that they reduplicated the words 'intuitively', and that each word required about 5 seconds or less for reduplication. All the participants used *-p, -m, -s,* and *-r* for the linker position, but none used *-A, -Il* or *-Am* as an additional infix. This result is shown in more detail in Table 1.

Looking closer at the results, most productions seem to satisfy the constraints previously reported (Demircan, 1987; Wedel, 1999; Wedel, 2000), but, certainly, some formations such as *böpböcek*, *fırfırın*, *mammasal* and *kemkemir*, violate these constraints. Unlike the study by Demircan (1987), the order of the linker type frequencies in this study is *-p* > *-s* > *-m* > *-r*. Moreover, the reduplicated forms with *r*-linkers seem to disconfirm Wedel's (1999; 2000) conclusion that *r*-forms might be lexicalized. They seem to be just less frequently used.

Explanations for these findings might lie in lexical statistics, in particular n-grams of graphemes, which we use in lieu of phonemes because of lack of speech data. For this end we used corpora to approximate a lexical statistic. When we examined the METU-Sabancı Turkish Treebank (Atalay et al., 2003), we found that there are 43,571 roots, of which 5,533 are distinct. The linker order found in the current study is exactly opposite of the frequency of words with roots that end

Table 1: TER results for the nonadjectives.

Word	*p*-linker	*m*-linker	*s*-linker	*r*-linker
bıçak	11	10	28	1
	(bıpbıçak)	*(bımbıçak)*	*(bısbıçak)*	*(bırbıçak)*
böcek	13	8	25	4
cevap	30	0	16	4
cami	28	8	14	0
çorba	29	0	16	5
dilek	32	11	7	0
davet	26	3	21	0
duvar	27	7	16	0
eğlen	50	0	0	0
fırın	17	2	23	8
felek	12	3	30	5
getir	32	0	18	0
götür	37	0	13	0
hüzün	43	2	5	0
jilet	36	6	8	0
kıble	18	2	30	0
kemir	23	8	14	5
leğen	36	3	11	0
laf	26	7	17	0
masal	14	5	29	2
nizam	32	8	10	0
pırasa	14	9	17	10
resim	38	4	8	0
surat	43	6	0	1
seçim	32	9	0	9
şerit	26	19	0	5
tutkal	37	5	8	0
tekerlek	24	7	18	1
vazo	29	2	19	0
yutkun	38	3	9	0
zarf	40	10	0	0
(%)	57%	11%	28%	4%

with the grapheme found in the linker; see Table 2, the second column. The third column repeats the same count in a much larger corpus, the BOUN corpus of 490 million words (Sak, Güngör, & Saraçlar, 2011). The database has 45,035 distinct stems, whose frequency of ending with *p/m/s/r* is reported in the third column. They have the same rank as the other corpus. We take the rank as an adequate relative measure of the lexical choice of endings in Turkish.

Next we consider the co-occurrence of the linker type with other consonants that are likely to be at the initial segment. For a word without TER beginning with the sequence $C_1V_1C_2$..., the consonant co-occurrence on the boundary of the "prefix" and the base will be one taken from the set $\{pC_1, mC_1, sC_1, rC_1\}$. Prefixation is a morphological operation, and we wanted to see if early lexical access can be contrasted with early morphological processing by frequency.

One hypothesis is that the linker may be selected so that the first segment of the reduplicated word has less resemblance to an existing root. One way to check this effect is to see if the consonant co-occurrence is minimized for the linker type-C_1 pairs. In order to test this hypothesis, the statistics from the METU Turkish Corpus (Say et al., 2002) is studied. Table 3 shows the number of distinct words containing the consonant co-occurrences composed of one linker and the initial grapheme of the nonadjectival word from the corpus. For example, 46 words in the corpus have *pb* as substring. Similarly, 482, 101 and 633 words have *pm*, *ps* and *pr* as substrings, respectively.

Table 2: Number of distinct stems/roots terminating with the same grapheme as a linker type, in two large corpora.

Root ending	METU-Sabancı	BOUN
p	100	639
s	128	780
m	281	1,620
r	470	3,523

Table 3: Linker-consonant co-occurrences in the corpus.

Consonant	Linkers			
	p-	*m-*	*s-*	*r-*
b	46	482	101	633
	(*pb*)	(*pm*)	(*ps*)	(*pr*)
c	44	435	136	705
ç	112	13	48	602
d	106	1599	148	9958
f	25	28	66	191
g	11	92	54	1519
h	189	114	168	200
j	7	1	7	90
k	275	134	845	2575
l	1799	3171	1655	8005
m	340	257	519	5156
n	90	82	140	559
p	100	447	404	499
r	952	201	139	277
s	529	926	612	3119
ş	10	90	10	624
t	820	109	4338	3321
v	25	25	61	61
y	132	122	719	346
z	36	161	16	195
(%)	6%	16%	17%	61%

The order of the linker type selection frequencies reported by the participants is exactly the opposite of the order of the frequency of the words in which the linker and the first consonant (C_1) co-occurred in the corpus. To exemplify: if the participants' choices occurred in the order $pC_1 > mC_1 > sC_1 > rC_1$ (where xC_1 indicates frequency of co-occurrence of x and C_1, in this sequence), then the linker type and conso-

nant co-occurrences in the corpus independent of reduplication are in the order $pC_1 < mC_1 < sC_1 < rC_1$. This seems to be true for all cases that we have tried. For example, the following are produced by the participants in varying frequencies for the word *masa* 'table': *mammasa, masmasa, mapmasa, marmasa*. When we look at the co-occurrence of *mm, sm, pm, rm* in the corpus, their frequency rank is the opposite of the ranking of the four alternatives by the participants. Such a direct relation to lexical frequency would be surprising if phonological and morphological contrasts were the sole bases of ranking as suggested by the constraints listed in (2). The same considerations apply to cophonology.

The Mann-Whitney non-parametric test was employed to compare the distributions of the two sets given in Table 1 and Table 3. It shows that the two sets are significantly different (U = 693, $p < .05$, $r = .79$). For the participants' responses, the frequency of answers significantly and negatively correlated with linker types, ranging from *-p* to *-r* (Pearson's $r(124) = -.64$, $p < .01$). On the other hand, the frequency of consonant collocations in the corpus significantly and positively correlate with linker types, ranging from *-p* to *-r* (Pearson's $r(124) = .34$, $p < .01$).

Discussion and Conclusion

The study of Turkish Emphatic Reduplication (TER) on unexpected targets, e.g. on nonadjectival words, and the 'intuitive' responses on part of the participants which violate the phonological feature constraints, show that although TER is a morphologically conditioned phonological process, it seems to depend on the knowledge of distributions such as consonant co-occurrence statistics and root ending statistics. The participants tended to dissimilate the linker type they chose from the known consonant co-occurrences and root endings of their language.

Besides the phonological constraints, selecting an appropriate linker so that the first segment of the reduplicated word has less resemblance to an existing root-word is additionally effective in the process of reduplication. This dissimilation tendency can be observed in Table 1. For example, *dar-davet*, *dur-duvar*, *göm-götür*, *gör-götür*, *ger-getir*, *gem-getir*, *hür-hüzün*, *var-vazo* and *zar-zarf* were not produced because *dar* 'tight', *dur* 'stop', *göm* 'bury', *gör* 'see', *ger* 'stretch', *gem* 'curb', *hür* 'free', *var* 'exist' and *zar* 'die' are already existing stems in Turkish. Thus, selecting a linker which has a frequent (admittedly orthographic) representation in the corpus would seem to steer the speaker to considering as if there was a root instead of a prefix. This we think might point out more ways to look at morphology-lexicon relation, rather than just the "blocking" kind such as *went/*goed* and *git/*gittir/götür* (leave/cause to leave/take away). Also, considering the fact that participants were able to respond within 5 seconds, it seems to us that the speakers are putting the co-occurrence frequencies in their language to online use.

The findings underline that cophonologies, i.e. morphological functions (Inkelas & Zoll, 2005), might require prior knowledge of known words' frequencies to select the appropriate linker after the truncation of the base in light of the phonological constraints. To be able to employ these statistics, TER as a process needs access to a speaker's lexicon.

We therefore suggest that Turkish emphatic reduplication, an apparently phonological operation, depends on global lexical knowledge for selecting an appropriate linker whose co-occurrence with the initial consonant of the reduplicated word is infrequent. (Yavas, 1980 was first to point out the lexical source of the linker type.)

We argue further that there are sufficient reasons to take emphatic reduplication as morpholexical, rather than phonological or cophonological. First, the ranking of TER elicitations from non-TER targets conforming to TER's base form paradigm is not consistent. This suggests that something other than phonological ranking is also at work.

The lexical constraints on TER seems to be more than the stem's part of speech and lexical statistics. The process is very productive when we can entertain a $((e,t),(e,t))$ reading for TER, from a $((e,t),(e,t))$ base, be it adjective or adverb, to $((e,t),(e,t))$ result. This is a semantic constraint. *Yemek* 'food' is not the right type (currently it seems to be (e,t)), and we have **yepyemek/*yemyemek/*yesyemek/*yeryemek*, although the first syllable of the base would be a phonologically legitimate input to TER if the semantic type could be satisfied: *yeşil* 'green' → *yemyeşil* 'all green'. Place names cannot be of the right semantic type either, and similarly fail to undergo TER if not forced: *Mordoğan* → **mosmordoğan*.

The semantics of reduplication works on aspectual properties or intensive aspects, depending on the morphological *and* lexical property of the $((e,t),(e,t))$ base: *çabuk* 'quick' → *çarçabuk* 'in haste', which is aspectual and derivational (cf. syntactic/phonological reduplication *çabuk çabuk* 'hurriedly'), and *mor* 'purple' → *mosmor* 'deep purple', which is intensive. Note also the case of *yeşil* → *yemyeşil* above, which seems to be intensive in some other way.

Connectives and postpositions are the hardest targets for TER, presumably because of their semantic type in addition to morphology: *ama* 'but' → **apama*, *göre* 'according to' → **göpgöre*. Other semantically potential targets for TER, e.g. VPs, are in fact (e,t), and as such they behave as expected: *uyu* 'sleep' → **upuyu*, *düşün* 'think' → **düpdüşün*. Additional linkers {*-A, -Il, -Am*} are never used by our participants, which suggests that such forms in (1) are probably lexicalized.

It is clear that the reduplicated "prefix", the linker types {*p, s, m, r*}, or the "infix" from {*-A, -Il, -Am*} are not morphological objects. They are not affixes or morphemes. It seems also clear that the process is not purely lexical or phonological. Its "allomorphy" is open-ended; there seems to be no discernible TER morpheme, or a purely morphophonological process.

We point out that (i) the process is codetermined by morphology and the lexicon, (ii) its semantics depend on lexical properties, and (iii) it cannot be repeated: *masmavi* → **masmasmavi* and *apaçık* → **apapaçık*. Therefore it is most likely

a morpholexical rule, with subsequent phonological effects, rather than causes, as exponence.

References

Atalay, N., Oflazer, K., & Say, B. (2003). The annotation process in the Turkish Treebank. In *Proceedings of the EACL workshop on linguistically interpreted corpora* (pp. 13–14). Budapest, Hungary.

Demircan, O. (1987). Emphatic reduplication in Turkish. In H. E. Boeschoten & L. T. Verhoeven (Eds.), *Studies on modern Turkish: Proceedings of the 3rd conference on Turkish linguistics* (pp. 24–41). Tilburg: Tilburg University Press.

Dhillon, R. (2009). Turkish emphatic reduplication: Balancing productive and lexicalized forms. *GLS*, *71*, 3–20.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. London: Routledge.

Güngör, T. (2003). Lexical and morphological statistics for Turkish. In *Proceedings of TAINN* (pp. 409–412).

Inkelas, S. (2000). Phonotactic blocking through structural immunity. In B. Stiebels & D. Wunderlich (Eds.), *Lexicon in focus* (pp. 7–40). Berlin: Akademia Verlag.

Inkelas, S. (2005). Morphological doubling theory: Evidence for morphological doubling in reduplication. In B. Hurch (Ed.), *Studies on reduplication* (pp. 63–86). Berling: Mouton de Gruyter.

Inkelas, S., & Orgun, C. O. (1995). Level ordering and economy in the lexical phonology of Turkish. *Language*, *71*, 763–793.

Inkelas, S., & Zoll, C. (2005). *Reduplication: Doubling in morphology*. Cambridge: Cambridge University Press.

Kelepir, M. (2000). To be or not to be faithful. In A. Göksel & C. Kerslake (Eds.), *Studies on Turkish and Turkic languages.* Harrassowitz Verlag.

Kılıç, O., & Bozşahin, C. (2012). Semi-supervised morpheme segmentation without morphological analysis. In *Proceedings of the workshop on language resources and technologies for Turkic languages, LREC.* İstanbul, Turkey: ELRA.

Kim, H. (2007). The full-to-partial reduction in Korean and Turkish reduplication. *Linguistic Research*, *26*, 121–148.

Kornfiltl, J. (1997). *Turkish*. London and New York: Routledge.

Lewis, G. (2000). *Turkish grammar*. Oxford: Oxford University Press. (2nd ed)

Orgun, C. O. (1999). Morphological doubling theory: Evidence for morphological doubling in reduplication. In B. Hermans & M. von Oostendorp (Eds.), *The derivational residue in phonological optimality theory* (pp. 247–267). Amsterdam: John Benjamins.

Oztaner, S. M. (1996). *A word grammar of Turkish with morphophonemic rules*. Master thesis, Department of Computer Engineering, Middle East Technical University.

Sak, H., Güngör, T., & Saraçlar, M. (2011). Resources for Turkish morphological processing. *Language Resources and Evaluation*, *45*(2), 249–261.

Say, B., Zeyrek, D., Oflazer, K., & Ozge, U. (2002). Development of a corpus and a treebank for present-day written Turkish. In K. Imer & G. Dogan (Eds.), *Proceedings of the eleventh international conference of Turkish linguistics* (pp. 183–192). Famagusta, Northern Cyprus: Eastern Mediterranean University Press.

Wedel, A. (1999). *Turkish emphatic reduplication*. Master thesis, Department of Linguistics, University of California, Santa Cruz.

Wedel, A. (2000). Perceptual distinctiveness in Turkish emphatic reduplication. In *Proceedingsof the 19th west coast conference on formal linguistics* (pp. 546–559).

Yavas, M. (1980). *Borrowing and its implications for Turkish phonology*. Unpublished doctoral dissertation, Univ. of Kansas.

Yu, A. (1999). *Dissimilation and allomorphy: The case of Turkish emphatic reduplication*. Master thesis, Department of Linguistics, University of California, Berkeley.