Storage and Computation of Multimorphemic Words in Turkish

Rabia Ergin (rabia.ergin@gmail.com)
Cognitive Science Program, Boğaziçi University

Emily Morgan (eimorgan@ucdavis.edu)
Department of Linguistics, University of California, Davis

Timothy O'Donnell (timothy.odonnell@mcgill.ca)
Department of Linguistics, McGill University

[Preprint version]

Abstract

Whether morphologically complex words are stored as a whole or decomposed into constituents has been well-investigated experimentally in Indo-European languages like English, Italian, Dutch and French. There is substantial evidence in these languages in favor of and architecture which allows both decomposition and storage. This study investigates how morphologically complex words involving two or more morphemes are represented in Turkish which, unlike Indo-European languages, is renowned for its highly rich morphology. Applying a probabilistic tradeoff-based model of morphological storage and computation (O'Donnell 2015) to a corpus of Turkish word forms, we derive predictions about stored patterns in the language. We discuss these patterns and propose several for future experimental investigation.

Keywords: Morphological processing; multimorphemic; frequency; decomposition; full-listing; Turkish

Introduction

Research into morphological processing has focused on how multimorphemic words are stored and accessed in long-term memory. Decompositional theories suggest that morphologically complex words are stored in terms of their component morphemes, with recognition occurring when all of a word's morphemes have been identified (e.g., Marslen-Wilson & Zhou, 1999; Taft & Forster, 1975). At the other end of the spectrum are full-listing theories, which suggest that morphologically complex words are stored as unanalyzed wholes in the mental lexicon (e.g., Butterworth, 1983; Manelis & Tharp, 1977; Seidenberg & Gomerman, 2000).

Between these two extremes are theories that posit that language users may store a mixture of morpheme-based and whole-word representations for multimorphemic words. Such approaches, often referred to as *dual-route*, *dual-mechanism*, or *dual representation* theories, propose that two different cognitive mechanisms, systems, or representations are involved in word structure—a decompositional route that analyzes words in terms of their component morphemes and a full-listing route that retrieves words as wholes from long-term memory. Under these accounts, a word may be processed using one or both routes depending on various properties such as whether it is a word or non-word, high or low frequency, derived or inflected, and regular or irregular (e.g. Baayen, Dijkstra & Schreuder, 1997; Burani & Caramazza, 1987; Cole, Beauvillain & Segui, 1989; Taft & Forster, 1975; Pinker, 1999; Jackendoff, 2002). For example, regular forms such as "liked" may follow a compositional route whereas the irregular forms such as "went" may follow a whole word route (e.g., Pinker, 1999).

Nearly four decades of research have suggested both subword and whole-word storage and computation (Amenta & Crepaldi, 2012; Lignos & Gorman, 2012). These data come from a variety of different paradigms (priming, lexical decision, eye-tracking, event-related potentials) and have been reported in a number of different languages (English, Dutch, French, Italian, Finnish). While it is still debated exactly what kinds of models most accurately capture the data, the wide scope of results and languages suggests that the involvement of both morpheme and whole-word representations is a robust phenomenon in multimorphemic word storage and computation. Such data have led to a number of theories of how such mixtures of computation and storage might be acquired by language learners, including the probabilistic tradeoff-based approach we study here (O'Donnell, 2015).

To date, however, the vast majority of the research on this topic has come from Indo-European languages that contain relatively simple morphological systems. Although these 'analytic' languages permit various morphological processes such as prefixation, suffixation, and compounding (in some cases with linking morphemes), the absolute complexity that words exhibit is rather limited. For example, although English words may contain multiple prefixes and suffixes (e.g., *non-de-compos-ition-al*), the English Lexicon Project ('ELP'; Balota, et al., 2007) reveals that English multimorphemic words contain only 2.5 morphemes on average. Furthermore, these languages contain relatively few multimorphemic words (the ELP lists a little over 50,000 for English) and many irregular and opaque forms, which may bias readers to rely on whole-word storage, even for regular/transparent forms. The consistency of the findings in favor of dual-route/dual-representation systems may thus be an artifact of the morphologically-impoverished languages that have been studied.

Only one morphologically *rich* (i.e., 'synthetic') language has been extensively studied from a psycholinguistic standpoint: Finnish. Finnish is a Uralic agglutinative language that allows words to be made from long strings of morphemes that would typically be expressed in Indo-European languages with sequences of words. The complexity of Finnish compared to most Indo-European languages is staggering—Karlson (1993; cited in Lehtonen & Laine, 2003), for example, suggests that each Finnish noun can have 2,000 different inflectional forms, made through different combinations of suffixes. In theory, the sheer size of the Finnish multimorphemic vocabulary may limit the number of forms that can be stored as wholes. In addition, the fact that Finnish is morphologically regular may bias the use of decomposition. As a result, Finnish is a valuable testing-ground for theories of morphological processing.

Despite the fact that Finnish has many properties that could bias readers to rely exclusively on decompositional processing, Laine and colleagues found evidence across a series of studies for both decompositional and whole-word processing (e.g. Lehtonen & Laine, 2003; Soveri, Lehtonen, & Laine, 2007). Although the results are not identical to Indo-European languages (specifically, whole-word representations were found only for the highest frequency Finnish inflected items whereas whole-word representations have been found for even lower frequency words in English, e.g., Alegre & Gordon, 1999; Lignos & Gorman, 2012), the fact that readers store both morphemes and words in a highly regular and complex language such as Finnish suggests that the dual-route/dual-representation architecture may be a fundamental component of multimorphemic word storage and computation rather than a fact that depends upon the morphological structure of a language.

Turkish

While the results from Finnish are intriguing, they represent only one language. It would be useful to have data from other morphologically rich languages—from other language families, ideally—in order to learn to what degree these other languages are characterized by mixtures of computation and storage.. Here we report the results of a prelimimary modeling study of Turkish multimorphemic words designed to make predictions about the pattern of storage and computation in this language. We anticipate the results of our study can be used in future experimental work designed to test the psychological reality of these claims with Turkish speakers.

Turkish is an Altaic language that, like Finnish, has an extremely productive morphological system, realized primarily through suffixation. The following complex words taken from the METU Turkish Corpus (Atalay, Oflazer, & Say, 2003) exemplify the morphological system. They are all legal words formed through successive suffixation:

(1) Göz – lük – lü[1]
Eye – DER – DER
"The one with the glasses"

(2) Güven – e – m – iyor – du – m
Trust – ABIL–NEG – PROG – PAST – 1stSg.
"I was not able to trust".

In the following complex example,15 suffixes combine to form a single word.

Kur: set up
Kur-um: institution
Kur-um-sal: institutional
Kur-um-sal-laş: become institutional
Kur-um-sal-laş-tır: institutionalize[2]
Kur-um-sal-laş-tır-ıver: institutionalize quickly
Kur-um-sal-laş-tır-ıver-e-me: cannot institutionalize quickly
Kur-um-sal-laş-tır-ıver-e-me-yecek: will not be able to institutionalize quickly
Kur-um-sal-laş-tır-ıver-e-me-yecek-ler: they will not be able to institutionalize quickly
Kur-um-sal-laş-tır-ıver-e-me-yecek-ler-imiz: those that we will not be able to institutionalize quickly
Kur-um-sal-laş-tır-ıver-e-me-yecek-ler-imiz-den: one of those that we will not be able to institutionalize quickly
Kur-um-sal-laş-tır-ıver-e-me-yecek-ler-imiz-den-miş: it has been heard/reported that (it) is one of those that we will not be able to institutionalize quickly
Kur-um-sal-laş-tır-ıver-e-me-yecek-ler-imiz-den-miş-siniz: it has been heard/reported that you are one of those that we will not be able to institutionalize quickly
Kur-um-sal-laş-tır[3]-ıver-e-me-yecek-ler-imiz-den-miş-siniz-ce-sine: it has been heard/reported that you are as if one of those that we will not be able to institutionalize quickly

Although Turkish suffixed words express ideas that are typically expressed as sentences in English, there are several reasons to believe that these items are in fact words, not sentences. First, other than the root, none of the suffixes can

[1] In these examples and elsewhere in this paper, we provide hyphens to delineate morpheme boundaries though these hyphens are not actually present in Turkish orthography..

[2] "Institutionalize" in this Turkish word means "to establish (something) as an institution", rather than "to put someone in an institution".

[3] All of these suffixes attached to a Turkish root have to agree with the preceding vowel in the string due to vowel harmony constraints. Vowel harmony is conditioned by the backness or frontness as well as roundedness or unroundedness of a vowel. As a result of the vowel harmony rules, almost every suffix in Turkish can have multiple forms. In addition, some initial consonants in a suffix undergo voicing assimilation with the preceding vowel or consonant (e.g., the suffix "–tır /tır/" has additional seven forms: –tir /tir/, –tur /tur/, –tür /tyr/,–dir /dir/, –dır /dır/, –dur /dur/, –dür /dyr/.

stand alone, indicating that they are all bound morphemes. Second, all of these suffixes yield to the constraints posed by Turkish vowel harmony rules (all vowels must agree in frontness/backness and high vowels must agree in rounding). These rules operate within but not across words.

Like in Finnish, Turkish morphology is highly productive—Hankamer (1989) estimates that the average multimorphemic word contains 4.8 morphemes. As in Finnish, Hankamer estimated that each verb can have over 2,000 inflectional forms—given a lexicon of 10,000 verb roots and 20,000 noun roots, Turkish could potentially contain over 200 billion lexical forms. Considering the huge number of entries and the highly regular nature of Turkish morphology, it has been claimed that "…the [Turkish] lexicon cannot be conceived as a stable, shared, inventory of words…" (Pierrehumbert, 2014), implying that Turkish speakers must rely on decomposition rather than storage for the majority of morphologically complex words.

To date, only one study has investigated the processing of Turkish multimorphemic words. Gürel (1999) tested monolingual Turkish speakers in a lexical decision task. The critical stimuli consisted of monomorphemic items such as *pencere* ('window') and inflected words containing one or two suffixes, such as *masa-da* ('on the table') and *oda-lar-dan* ('from the rooms'). The suffixes used were the plural marker *–lEr*, the locative *–DE*, and the ablative *–DEn*.[4] Of these three suffixes, the plural is the most frequent, followed by the locative and then the ablative. Mono- and multimorphemic words were matched on stem frequency but not word frequency since these values were not available at the time the study was conducted.

Gürel found a significant effect of suffix frequency. While bimorphemic words containing the ablative (least frequent) suffix had significantly longer RTs than the monomorphemic words, words containing the locative and plural suffixes did not differ significantly from the monomorphemic words. Moreover, words containing two suffixes (plural + ablative, plural + locative) were responded to just as quickly as words containing just one suffix (e.g., ablative, locative). Gürel concluded that not all multimorphemic words in Turkish are accessed through decomposition. In particular, words inflected with a single frequent suffix seem to be accessed without any decomposition. She also proposed that in words with more than one suffix, the stem+plural combination may be retrieved as a whole, perhaps being combined with the word-final suffix compositionally. While these results suggest that at least some holistic processing occurs in Turkish, the results come primarily from the plural suffix, which limits the generalizability of the conclusions. Additionally, the stimuli were not controlled for word frequency, leaving open the possibility that other factors were involved in driving the observed effects.

[4] Capital letters are used indicate graphemes and phonemes that alternate (vary) in accordance with Turkish vowel harmony and voicing assimilation rules.

Present Investigation

In the present study, we apply a probabilistic *tradeoff-based* model of morphological storage and computation – known as fragment grammars (O'Donnell, 2015)—to make predictions about which combinations of Turkish morphemes might be stored and which computed. We discuss our findings below and provide suggestions for structures which can be studied in future experimental work on Turkish morphological processing.

The tradeoff-based approach is designed to distinguish freely combining productive units in a language (e.g., words and morphemes) from recurring patterns which do not generalize, but are rather stored together within larger structures. It does this by optimizing a balance between two competing biases. The first bias favors smaller more compact lexicons with highly reusable units. The second favors simple derivations of individual forms, involving fewer lexical items. These two biases are opposing–if units are smaller in general, the lexicon will contain few items, but more units will be needed to derive individual forms. On the other hand, if units are larger then forms can be derived using fewer steps, but the lexicon will have to contain many, less reusable forms. By optimizing this tradeoff on a particular input dataset, the model makes specific predictions about the pattern of computation and reuse in a language. The tradeoff-based approach has been used successfully in a number of morphological systems, but has not been applied to a morphologically rich agglutinative language such as Turkish.

When applied to an the model produces predictions about stored tree fragments such as those shown in Figure 1. On the left hand side of the figure is a tree fragment which stores all surface morphemes down to leaves (tarafından "by"). On the right hand side is a tree fragment which predicts that the stored combination of morphemes marking the third person possessive and ablative case ("by his/her N") can freely combine with an arbitrary root variable of category Noun.

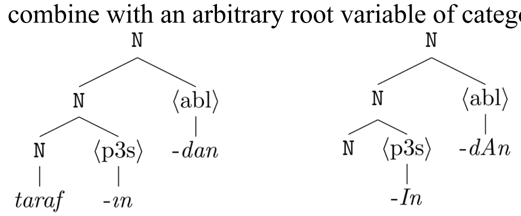

Figure 1. Two tree fragments discovered by our model. The fragment on the right represents a fully stored word while the fragment on the right stores a possessive marker and case ending together with a generic stem.

Fragment Grammar and similar approaches (Bod, Scha, & Sima'an, 2003; Johnson et al., 2007; Post et al. 2013) provide a way of bridging generative and constructional (Goldberg, 2003; Dabrowska, 2014) approaches to language since they explain how complex constructions can be built and stored out of simpler parts. In this study, we train the model on a corpus of Turkish words. The model predicts specific stored fragments of structure (together with their probabilities); we discuss these predictions.

Methods

Corpus construction We began with the METU-Sabancı Turkish Treebank (Atalay, Oflazer, & Say, 2003), a corpus of 173,469 Turkish word types. We obtain initial morphological parses for all words in the corpus using the TRmorph morphological analyzer and disambiguator (Çöltekin, 2014). For the majority of word types in the corpus, this disambiguator assigns a single best morphological analysis, which we take as the single analysis for our corpus.

For 33,152 word types, TRmorph cannot assign a best analysis. For these words, we begin by considering all possible analyses according to TRmorph (a total of 253,337 analyses across all word types). From there, we hand-constructed a set of heuristics to narrow the list of possible analyses to only those that were truly possible in Turkish. 54,624 total analyses for 22,963 word types passed our heuristic filters—that is on average, each of these word types had 2-3 possible analyses in our corpus. We included all such analyses in our corpus, distributing token counts obtained from the METU corpus evenly over word types.

Fragment Grammar analysis We ran the fragment grammars stochastic search algorithm through the corpus for 1000 sweeps. Parameters were set to Pitman-Yor discount parameter was set to 0, the Pitman-Yor concentration parameter was set to 1, and all pseudocounts were set to 1 for all outcomes on all Dirichlet-Multinomial distributions (see, O'Donnell, 2015 for details on the meaning of parameter values). We performed all following analyses on the best scoring grammar output by the model.

Evaluation

To evaluate the Fragment Grammar, we examined the probable predicted nominal and adjectival tree fragments (500 of each) that contained at least two leaves at the frontier of the fragment. That is, we examined all tree fragments that stored two or more surface morphemes together with any number of categorial variable slots for other morphemes.

Nouns

Of the 500 combinations, we were able to categorize 482 into one of the patterns listed in Table 1. The remaining 18 combinations were corpus or analytical errors that could not be categorized for one of the following reasons: (i) the root was not a meaningful Turkish word, (ii) the root was not parsed into its further constituents or (iii) the word constituted an illegal combination of inflectional morphemes. We exclude these forms from our analyses.

Table 1 lists the patterns in the top 500 nominal combinations predicted by the model. In these patterns, N/V/A refer to nominal, verbal, or adjectival root which could appear either instantiated as a specific root or as a variable (allowing any root of the appropriate category) while POSS, CASE, etc. always refer to specific inflectional endings (e.g, the first person possessive marker -m as in "masa-m"; *Eng*., my table). The patterns are further annotated for the number of combinations involving a specific root (414) or a generic variable in root position (68). 13 combinations, categorized as Other, appear once in the list with no detectable patterns common with any other combinations.

Pattern 1: N = N + POSS – a nominal root followed by a possessive maker (e.g., bakanlığ-ı in (3)) is the most common combination and all of the combinations having this pattern involve a specific root. The pattern with the third singular possessive marker, that is root + <i>, is a highly plausible candidate for storage because of the high frequency of noun-noun compounds in which root + <i> appears in second position. An example is shown in 3.

Table 1. Noun combinations

Pattern	# of combinations with specific root	with generic root
1. N = N + POSS	112	0
2. N = N + POSS + CASE	69	17
3. N = N + CASE	85	0
4. N = V + INF	39	0
5. N = N + PL	23	0
6. N = N + PL + POSS	20	3
7. N = V + vn:past/fut + POSS + CASE	11	6
8. N = N + PL + CASE	8	6
9. N = V + INF + POSS	10	2
10. N = N + PL + POSS + CASE	2	8
11. N = V + INF + CASE	7	3
12. N = V + TENSE + POSS	4	2
13. N = N + cpl-past/pres /evid/cond + PERS	0	6
14. N = V + vn:past/fut + PL + POSS + CASE	1	4
15. N = N + LIK	5	0
16. N = V + <yis>	5	0
17. N = N + LIK + POSS	2	2
18. N = V + INF + PL	3	0
19. N = N + CI	2	0
20. N = N + CI + LIK	1	1
21. N = Other	5	8

(3) Eğitim Bakanlığ – ı
Education Ministry – POSS
"Ministry of Education"

Similarly, the model predicts N + <m> – a noun with a first person possessive markers–a strong candidate for storage. It is a combination frequently used in isolation while addressing people in certain contexts:

(4) Tabi, efendi – m
Yes, master/sir – POSS
"Yes, (my) master/sir."

Pattern 2: The second most frequent pattern is N = N + POSS + CASE, a case-marked noun with a possessive marker (e,g., ev-im-de "HOME-p1s-loc" *Eng*. at my home). Many of the stored examples consist of locatively marked, high-frequency roots, (e.g., ev-im-de “HOME-p1s-loc” *Eng*. at my home; ev-in-den “HOME-p2s-abl”, *Eng*. from your home). Furthermore, N + POSS + CASE combinations involving specific roots, such as the ones listed below, are plausible candidates for storage because the whole forms, while morphologically nominal, are frequently used as postpositions in Tukish.

(5) N = iç <p3s> <loc>
iç – in – de (*Eng*., in)
N = taraf <p3s> <abl>
taraf – ın – dan (*Eng*., by)
N = neden <p3s> <ins>
neden – i – yle (*Eng*., because of)
N = alt <p3s> <loc>
alt – ı – nda (*Eng*. under)
N = yıl <p3s> <loc>
yıl – ı – nda (*Eng*. at the year of)

Pattern 3: This pattern, N = N + CASE, consists of stereotyped case marked nouns (e.g., durum <loc>, durum – da, “under the circumstance”). Common combinations in the model output are exemplified below, (6-8).

(6) N = durum <loc>
Bu durum – da
This circumstance – LOC
“Under this circumstance”

(7) N = neden <ins>
Bu neden – le
This reason – with
“Because of this/As a result of this”

(8) N = dikkat <dat>
Konu–yu dikkate–e al – ma – dı.
Issue–ACC account–DAT take–NEG–PAST
“(He) did not take the issue into account.”

Some further observations: Many patterns, such as patterns 1, 3, 4, 5, 15, 16, 18 and 19 occur only with specific roots rather than with generic root variables. Longer combinations tend to involve fewer specific roots (e.g., patterns 10, 13 and 14) and instead involve a root variable. 18 out of 482 combinations involve derivational morphology (i.e., Patterns 15-20).

Adjectives

406 out of 500 combinations are categorized into patterns listed in Table 1. 94 combinations cannot be categorized for a pattern in the adjective list for the following reasons: 1) the root is not a meaningful Turkish word (9 of them), 2) the whole form is not an Adjective (77 of them), 3) illegal combination of inflectional morphology (8 of them). 11 combinations, categorized as Other, appear once in the list with no detectable patterns in common with any other combinations. 141 out of 406 of the combinations in the adjective list involve derivational morphology (i.e., Patterns 2, 4, 6, 7 and 12). Other patterns are adjectival clauses. As opposed to the noun combinations, adjective combinations involve fewer generic roots (14.3% *vs*. 3.2%, respectively).

Table 2. Adjective combinations

Pattern	# of combinations with specific root	with generic root
1. Adj = V + <part:pres>	118	0
2. Adj = N + LI	96	0
3. Adj = V+ part:past/fut+POSS	79	7
4. Adj = N + SAL	18	0
5. Adj = N + POSS + LOC + KI	17	1
6. Adj = N + SIZ	17	0
7. Adj = N + LIK	9	0
8. Adj = Adv + KI	9	0
9. Adj = V/N+lAn+part:press/fut	9	0
10. Adj = N + LOC + KI	6	1
11. Adj = V <neg> <part:pres>	3	1
12. Adj = Adj + lAs + part:pres	1	1
13. Adj = Other	9	2

Pattern 1: Adj = V + <part:pres> is the most frequent combination in the adjective list.

(9) Adj = yap <part:pres>
yap – an
“the one who is doing”

(10) Adj = git <part:pres>
gid – en
“the one who is going”

Pattern 2: Adj = N + LI is the second most frequent combination in the adjective list.

(11) Adj = önem
önem – li
(*Eng*. importance → important)

(12) Adj = çeşit
çeşit – li
(*Eng*. variety → various)

LI turns nouns into Adjectives meaning “involving” the content of the noun. For example, önem means ‘importance’ and önemli means “important” –involving importance. Likewise, çeşit means “variety” and çeşitli means “various” –involving variety. LI in pattern 2 and also –sIz in pattern 6 below are highly productive suffixes allowing the creation of novel forms on the spot. There are many high frequency adjectives marked with –LI and –sIz which are good candidates for storage. On the other hand, novel forms created with –LI and –sIz on the spot are good candidates for decomposition as in (13) and (14).

(13) Timsiz parti eğlenceli değildi.
Tim-without party fun not
“Party without Tim was not fun”

(14) Bir daha Timli parti yapalım
Next time Tim-with party have-will
“We will have a party involving Tim next time”

Pattern 3: Adj = V+part:past/fut+POSS is the third most frequent pattern.

(15) Adj = V <en> <i>
gel – en – i
(*Eng*. the one who comes)

Pattern 4: Adj = N + SAL. This pattern derives an adjective from a noun with the suffix –sAl.

(16) Adj = din <sal>

dinsel
(*Eng*. religion → religious)

Pattern 6: Adj = N + SIZ. This pattern derives an adjective from a noun with the suffix –sIz (*Eng*. –less, without).

(17) Adj = son <siz>
son – suz
(Eng. end → eternal)

(18) Adj = çare <siz>
care – siz
(Eng. solution → desperate)

Pattern 7: Adj = N + LIK. –lIk is a highly productive suffix in Turkish.

(19) Adj = lira <lik>
lira – lık
(*Eng*. worth of [X amount of] Lira)

Similar to the suffix – LI in Pattern 2, -sAl in Pattern 4, –sIz in Pattern 6 and –lIk in Pattern 7 are highly productive suffixes. Frequent surface forms are good candidates for storage whereas infrequent novel forms are good candidates for decomposition.

Conclusions

We have reported the results of a preliminary study deriving predictions for storage and computation of word forms in Turkish – a morphologically rich language. Despite the fact that Turkish is highly regular, we found that the probabilistic tradeoff-based model of O'Donnell (2015) predicted a number of patterns of nouns and adjectives which were plausible candidates for storage. We highlighted the comparison between patterns which contained a stored root and those which contain a generic root variable. These particular cases can form the basis for experimental manipulations testing the psychological reality of our claims.

References

Atalay, N. B., Oflazer, K., & Say, B. (2003). The Annotation Process in the Turkish Treebank. Presented at the Proceedings of the EACL Workshop on Linguistically Interpreted Corpora - LINC, Budapest, Hungary.

Amenta, S. & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Language Sciences*, 3, 232.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: evidence for a parallel dual-route model. Journal of Memory and Language, 37, 94–117.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior research methods*, *39*(3), 445-459.

Bod, R., Scha, R., & Sima'an, K. (Eds.). (2003). *Data-oriented parsing*. University of Chicago Press.

Burani, C., & Caramazza, A. (1987). Representation and processing of derived words. *Language and Cognitive Processes,* 2, 217–227.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production* (Vol. 2). New York: Academic Press.

Colé, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. Journal of *Memory and language, 28*(1), 1-13.

Çöltekin, Ç. (2014). A Set of Open Source Tools for Turkish Natural Language Processing (pp. 1079–1086). Presented at the In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).

Dabrowska, E. (2014). Recycling utterances: A speaker's guide to sentence processing. Cognitive Linguistics, 25 (4), 617-653.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5), 219–224.

Gürel, A. (1999). Decomposition: To what extent? The case of Turkish. *Brain and language*, *68*(1-2), 218-224.

Hankamer, J. (1989). Morphological parsing and the lexicon. In *Lexical representation and process* (pp. 392-408). MIT Press.

Jackendoff, R. S. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Johnson, M., Griffiths, T. L., & Goldwater, S. (2007). Adaptor grammars: A framework for specifying compositional nonparametric Bayesian models. Advances in Neural Information Processing Systems, 19, 641–648.

Lehtonen, M., & Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and cognition*, 6(3), 213-225.

Lignos, C., & Gorman, K. (2012). Revisiting frequency and storage in morphological processing. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* (Vol. 48, No. 1, pp. 447-461).

Manelis, L., & Tharp, D. A. (1977). The processing of affixed words. *Memory and Cognition,* 5, 690-695.

Marslen-Wilson, W., & Zhou, X. (1999). Abstractness, allomorphy, and lexical architecture. *Language and Cognitive Processes*, 14, 321-352.

Niemi, J., Laine, M., & Tuominen, J. (1994) Cognitive morphology in Finnish: Foundations of a new model. *Language and Cognitive Processes* 9: 423–446.

O'Donnell, T. (2015). *Productivity and Reuse in Language: A Theory of Linguistic Computation and Storage.* MIT Press.

Pierrehumbert, J. (2014). Real Words, Possible Words, and New Words. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. New York: Perennial.

Post, M. and Gildea, D. (2013). Bayesian tree substitution grammars as a usage- based approach. *Language and Speech, 56*(3):291–308.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, 4, 353-361.

Soveri, A., Lehtonen, M., & Laine, M. J. (2007). Word frequency and morphological processing in Finnish revisited. *The mental lexicon*, 2(3), 359-385.

Taft, M., & Forster, K.I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14, 638-647.