2023. Proceedings of the Workshop on Turkic and Languages in Contact with Turkic 8. 103–117.

Binomial expressions as a window on the phonological system of Turkish

Serra Gök & Stefano Canalis*

Abstract. The principles governing word order in binomial expressions have been investigated in a few languages, revealing the relevance of several interacting semantic and phonological constraints, some of them universal or near-universal while others appear to be more language-specific. This paper presents the results of a preliminary quantitative analysis of some constraints influencing word order in Turkish irreversible binomials. We found that semantic constraints referring to the semantic prominence and iconic order of the items of the binomial, a constraint placing loanwords in final position, suprasegmental phonological constraints referring to syllable count, presence of onsets and presence of codas, segmental phonological constraints referring to the sonority of consonants, vowel roundness and presence of an initial labial consonant, all influence word order in Turkish irreversible binomials. Furthermore, the evidence from the segmental constraints may also shed light on some aspects of the Turkish phonological system, i.e. the markedness of rounded vowels and labial consonants, the properties of the sonority scale in Turkish and the featural specification of Turkish /j/.

Keywords. binomial expressions; Turkish phonology; word order; linearization; semantic constraints; phonological constraints; sonority scale

1. Introduction. The present study is a preliminary quantitative investigation of word order in Turkish irreversible binomial expressions, with a focus on segmental phonological factors and the implications they have for our understanding of the phonological system of Turkish.

Research on the principles governing word order in binomial expressions (which we define, following Malkiel (1959: 113), as "sequence[s] of two words pertaining to the same form-class, placed on an identical level of syntactic hierarchy, and ordinarily connected by some kind of lexical link" and often but not necessarily having an idiomatic meaning; see section 2 for a discussion of the validity of this definition for the Turkish constructions analyzed here) has a long history. In fact one of them, namely the tendency for the word with most syllables to be final, was apparently first formulated by the ancient Indian grammarian Pāṇini (Cooper & Ross 1975: 78). When both orders are attested (that is, when the binomial is 'reversible', again adopting Malkiel's terminology), a preference for a certain order over the other is usually clearly visible. For example, while both *demand and supply* and *supply and demand* are widely attested in English texts, the latter is much more frequent than the former, their ratio being roughly six to one in 2019 according to Google Ngram Viewer (see Benor and Levy 2006 for a detailed survey on the constraints contributing to ordering preferences in English reversible binomials; they provide a combination of twenty semantic, metrical phonological, non-metrical phonological, frequency-based and alphabetic constraints influencing the word order of English binomials, which gives an indication of the complexity and heterogeneity of the principles that may be relevant in a given language).

* We would like to thank the audience at TU+ 8, two anonymous reviewers, Furkan Dikmen, Muhammed İleri and Marinela Sotiri for their comments and suggestions. Authors: Serra Gök, Boğaziçi University (serra.gok@hotmail.com) & Stefano Canalis, Boğaziçi University (stefano.canalis@boun.edu.tr).

Other binomials have an 'irreversible' (also known as 'frozen') word order (e.g. English *odds and ends*); the inversion of the canonical order would be practically unacceptable. Interestingly, although the rigidity of their word order implies they are lexicalized constructions, this does not mean it is arbitrary; the same factors active in reversible binomials can be observed in irreversible ones (in this case, we observe more binomial types that conform to a certain constraint than types that do not, rather than more binomial tokens of a given binomial type conforming to a certain constraint than tokens not doing).

At the same time, while binomial constructions are quite common cross-linguistically, they have been investigated in some detail in only a small number of languages. Some parameters, especially among the semantic ones, may have a universal or near-universal validity (for instance, the word order in *wait and see* reflects the temporal order of the two actions), but others may be more language-specific, and relatively little is known about the factors governing word order in Turkish binomials. Furthermore, quantitative measurements of the validity of the constraints proposed for Turkish are limited, the only preliminary attempt being Tuna (1982-1983). The present study aims to provide some exploratory insights on the word order of Turkish frozen binomials, especially regarding the phonological constraints that refer to vocalic and consonantal features.

The structure of this paper is outlined as follows: we examine some general properties of Turkish binomials in the next section. We discuss some factors that have been observed to influence word order in binomials cross-linguistically in section 3. We provide a summary of the constraints proposed in the previous literature on word order in Turkish binomials in section 4, while in section 5 we present the criteria we adopt. We describe our dataset in section 6, and in section 7 we examine the effectiveness of our criteria in predicting word order in Turkish binomials, and briefly comment our results. Section 8 is devoted to a more detailed discussion of the segmental phonological constraints that pertain to vocalic and consonantal phonological features. Finally, we present our general conclusions in section 9.

2. What is a binomial expression in Turkish? Building on Yakov Malkiel's definition quoted above, we consider irreversible binomials in Turkish to be lexicalized fixed sequences of two (or occasionally more) words belonging to the same word class, which may or may not be connected by some kind of lexical link, and often have a non-compositional meaning; some examples are given in (1).

(1)	*acı tatlı*	'bitter – sweet'[1]
	kan-lı can-lı	'blood-with – soul-with' / 'alive'
	it köpek	'dog – dog' / 'rascal'
	(bir) deri (bir) kemik	'skin – bone' / 'very skinny'
	kurt kuş	'wolf – bird' / 'dangerous people'
	çayır çimen	'meadow – grass' / 'meadows'

[1] Hereafter, we adopt the following notational conventions. Following Malkiel (1959: 114) and Benor & Levy (2006: 235), we call the first position of a binomial *slot A* and the second position *slot B*. The words themselves will be called *item A* and *item B*, or, when there is no risk of confusion, simply *A* and *B*. An input binomial is an unordered pair of words {X, Y} (which will be listed in alphabetical order), which in principle could be linearized as either XY or YX; a surface binomial is an actual ordered form AB. Thus, the surface form of the input binomial {*acı*, *tatlı*} is *acı tatlı*, where *acı* is in slot A, and *tatlı* is in slot B. When the meaning of the binomial is compositional, we report the lexical meaning of each word separated by a dash. When the meaning of the binomial is not compositional, lexical meanings are followed by the idiomatic meaning of the whole expression preceded by a slash. Thus, the gloss of *acı tatlı* is 'bitter – sweet', while the gloss of *iki üç* is 'two – three' / 'a small number, few'.

Three aspects of this definition may deserve a closer look, i.e. the frequent absence of a lexical conjunction in Turkish binomials, their morphological or syntactic nature, and the relationship of our definition with the notion of *ikileme* (which can be translated as 'reduplication' or 'doubling') traditionally used in Turkish linguistics.

2.1. LEXICAL LINK. Malkiel (1959: 113) observed that in most languages binomial items are "ordinarily connected by some kind of lexical link", typically either a conjunction or a preposition, such as[2] English *odds and ends*, *(come) hell or high water*, *side by side*, German *Sturm und Drang*, modern Hebrew *nisim ve niflaot* 'miracles and wonders' (Handelsmana & Bat-Elit 2022), which most often but not necessarily occurs between A and B (cf. the conjunction *-que* in Latin *terrā marīque* 'by land and sea'). However, the presence of a lexical link is not a necessary property of binomials (cf. English *upside down*). What is necessary is that the words belong to the same word class, and, in the case of frozen binomials, have fixed order; additionally, the meaning of many binomials is lexicalized. Thus, we believe that the absence of any overt link between the two lexical items of most of the Turkish constructions we discuss below is not an impediment to classify them as *bona fide* binomials.

2.2. COMPOUNDS OR PHRASES? The absence of a link in many Turkish binomial constructions also calls into question their status as (lexicalized) coordinated phrases. If two lexical words occur together without any linking element, it might be argued they form a compound word rather than a syntactic phrase, especially if the meaning of the construction is idiomatic. In fact, English *bittersweet*, semantically and structurally similar to Turkish *acı tatlı*, is routinely assumed to be a compound, not a phrase, and at least one work (Arcodia & Sarı 2018) considers expressions such as *acı tatlı* to be coordinating compounds, or more precisely "conventionalised/lexicalised coordinating constructions located between morphology and syntax [which] are not prototypical morphological compounds […], but [….] show some degree of conventionalisation and fixity, differently from phrases" (Arcodia & Sarı 2018: 336). We believe it is more appropriate to consider them irreversible binomials, i.e. lexicalized coordinated phrases with a frozen word order. First, when inflectional markers are present, they occur in both roots (e.g. *önün-de sonun-da* 'front-LOC – end-LOC' / 'sooner or later', *var-a yoğ-a* 'existent-DAT – absent-DAT' / 'anything', *ol-du bit-ti* 'happen-PFV – to be over-PFV' / 'fait accompli'), whereas compounds typically show lexical integrity. Second, uncontroversial compounds in Turkish are not subject to the phonological and semantic constraints we examine below. Third, some constructions without an overt conjunction have a parallel (and clearly phrasal) conjoined counterpart (cf. *gece gündüz* 'night' – 'day' / 'nonstop' and *gece ve gündüz* 'night and day'). Finally, we wholeheartedly agree that they show "some degree of conventionalisation and fixity"; however, these are telltale signs of binomial constructions, especially in irreversible binomials, which often have a non-idiomatic meaning and by definition have a fixed word order. Thus, we assume here that constructions such as *acı tatlı* are (lexicalised) phrases rather than compounds. However, it is worth adding that this assumption is not a crucial aspect of the account of word order we present below. Since we are mainly concerned with the principles governing the order of the two items within these constructions, rather than on the morphological/syntactic status of the constructions themselves, our findings would remain relevant in either case.

2.3. BINOMIALS AND *IKILEMELER*. The last aspect of our definition that may benefit from a comment is its relationship with the category of *ikileme*, frequently employed in Turkish linguistics

[2] Examples from Malkiel (1959) unless otherwise specified.

(Göksel & Kerslake 2010; see also Ağakay 1953, 1954, Hatiboğlu 1981, Tuna 1982-1983, Çoraklı 2001) to refer to morphological or syntactic processes that partially or completely copy a word or phrase. While most binomial constructions discussed here belong to both categories, the two are not perfectly synonymous. On the one hand, three types constructions traditionally included in the category of *ikileme* are not binomials, and will not be discussed in the following sections. The first type is a prefixation process that semantically intensifies the quality of the root (usually an adjective), and phonologically is a partial copy of the initial segments of the root (e.g. *uzun* 'long' → *up-uzun* 'very long'). The second is so-called *m*-reduplication, a morphological process that "generalize[s] the concept denoted by a particular word or phrase to include other similar objects, events or states of affairs" (Göksel & Kerslake 2010: 99), and is realized doubling the word or phrase and adding an initial /m/ if the word/phrase begins with a vowel, or turning the initial consonant to /m/ if the word/phrase begins with a consonant: *etek metek* 'skirt(s) and the like' from *etek* 'skirt', *çay may* 'tea and the like' from *çay* 'tea'. The third is full reduplication (e.g. *yavaş yavaş* 'slowly' from *yavaş* 'slow'). The first process being affixation, clearly it is not a binomial construction. The latter two might be considered as two peculiar types of binomials, but there is very little (in the case of *m*-reduplication) or nothing (in the case of full reduplication) to say about their word order, which is the subject of the following sections.

On the other hand, some binomials are not *ikileme*, if (partial or total) reduplication is assumed to be the defining property of the latter. It is hard to consider constructions such as *al mor* 'red – purple' / 'colorful, ashamed', where no segment of A is present in B, as instances of reduplication; yet, they meet the standard definition of binomial, and the same constraints that govern their word order also govern word order in Turkish constructions where A and B are phonologically similar. Therefore, we believe it is appropriate to include both partially reduplicating and non-reduplicating constructions within the category of Turkish binomials. In this respect, it is worth noting that several works investigating *ikilemeler* (see e.g. Hatiboğlu 1981, Tuna 1982-1983) include constructions such as *al mor* in their examples, implicitly admitting that lack of reduplication is not sufficient to classify them into a separate category.

This is not to deny that a peculiarity of Turkish binomials, compared to binomials in other languages, is the very frequent phonological similarity of the two words, which is why they are traditionally counted among *ikilemeler*. They often fully rhyme (e.g. *çat pat* 'onomatopoeic word – onomatopoeic word' / 'broken speech') or half rhyme, usually sharing the final consonant (e.g. *az buz* 'little – ice' / 'a little'). Sometimes the whole final syllable or everything save for the initial consonant is identical (e.g. *apar topar* 'meaningless word – meaningless word' / 'in a rush', *açık saçık* 'open – spread' / 'obscene'), sometimes only vowels change (e.g. *kem küm* 'onomatopoeic word – onomatopoeic word' / 'stammering' and *zar zor* 'onomatopoeic word – difficult' / 'barely'), and often initial consonants are identical (e.g. *taş toprak* 'stone – soil'). Regarding their semantics, Turkish binomials may include meaningless words that only exist as part of a given binomial, a possibility attested in other languages (cf. English *spick and span*), but apparently exploited more frequently in Turkish; the meaningless word is usually in slot B (e.g. *eski püskü* 'old – tattered' / 'ragged', *bölük pörçük* 'company, squad – meaningless word' / 'piecemeal').

We will restrict our analysis below to binomials having a frozen word order. This choice was made because a) most Turkish binomial have fixed order, b) at least for non-idiomatic constructions, the boundary between reversible binomials and 'true' phrases may be difficult to delimit, and c) our current data consists of a list of (irreversible) binomials; reversible binomials

would also require the inspection of a large[3] corpus of Turkish texts, as any investigation to establish the relative frequency of AB and BA tokens for each reversible binomial type should be based on frequency data. Nevertheless, we suppose the criteria accounting for word order in frozen binomials may be broadly valid for reversible binomials as well, and we plan to investigate them in the future.

3. Factors frequently affecting word order in binomials. The literature on the factors that have been found to influence word order in binomials is extensive (among many others, Jespersen 1905, Malkiel 1959, Bolinger 1962, Cooper & Ross 1975, Birdsong 1979, Oakeshott-Taylor 1984, Benor and Levy 2006, Mollin 2014, Green & Birdsong 2018 on English, Jacobsen 1982 on Basque, Oakeshott-Taylor 1984 on Afrikaans, Lambrecht 1984, Oakeshott-Taylor 1984, Müller 1997 on German, Masini 2006 on Italian, Kaye 2009 on Arabic), even if unevenly distributed across languages and language families, as our list shows. Therefore, here we limit ourselves to a brief discussion of the most important categories of constraints proposed for languages other than Turkish.

The role of suprasegmental properties in determining word order in binomials has been noticed long ago. As mentioned in section 1, the tendency for words with more syllables to be in slot B was already identified by Pāṇini. The internal structure of syllables may be relevant too, the preferred position of more complex subsyllabic constituents being slot B. Onsets and codas tend to have fewer consonants in slot A than in B (Cooper & Ross 1975, Green & Birdsong 2018), while long vowels tend to occur more often in B (Bolinger 1962, Cooper & Ross 1975, Green & Birdsong 2018). Another suprasegmental property, i.e. stress, often plays a role. Already Jespersen (1905: 232-233) noted that English binomials tend to avoid stress lapses, favoring a binary rhythm ˈσσˈσσ (such as in *bread and butter*) over sequences of two unstressed syllables ˈσσσˈσ (**butter and bread*). Binomials also tend to avoid final stress on item B (Bolinger 1962 with regard to English, Müller 1997 with regard to German).

Along with metrical phonological properties, segmental phonological properties often play a role. Back and/or rounded vowels tend to be placed in slot B (Cooper & Ross 1975, Oakeshott-Taylor 1984, Green & Birdsong 2018), while stressed vowels tend to be higher in slot A than B (Birdsong 1979, Green & Birdsong 2018). With regard to consonants, sonority is probably the most well-studied factor (see especially Cooper & Ross 1975, Green & Birdsong 2018). Word-initial consonants should be higher on the sonority scale in slot A than B (*wear and tear*), while the opposite holds for word-final consonants, which tend to be more obstruent in slot A than B (*safe and sane*).

It might be said that many[4] phonological constraints show a preference for the least marked option (lower rather than higher syllable count, short rather than long vowel, unrounded rather than rounded vowel, and so on) to be in slot A. A is the unmarked or less marked item for semantic constraints as well. For instance, items having a broader, more general meaning and a greater freedom of distribution tend to occur first (e.g. *pull and tug*, *flowers and roses*, Benor & Levy 2006). So also do morphologically simpler items, as in *complete and unabridged*, where only item B is polymorphemic. Alongside these formal types of markedness, another kind of

[3] The occurrence frequency of most binomials is fairly low, which requires a large amount of data to be able to observe ordering preferences.

[4] Although not all. For example, onsetless syllables are supposedly more marked than syllables with an onset (among many others, Gordon 2006: 85), yet the preference for a lower number of initial consonants in slot A not only favours simple onsets over clusters (*make or break*), but also initial onsetless syllables over syllables with an onset (*eeny-meeny-miney-moe*).

markedness that is often relevant in binomials is what Benor & Levy (2006) call 'perceptual' markedness: an asymmetric relationship between the items of a binomial that can be recognized only through extralinguistic, real-world knowledge, often reflecting perceptual or cultural biases. Thus, animate precedes inanimate (*deer and trees*), proximal precedes distal (*here and there*), male precedes female (*husband and wife*; see also Wright et al. 2005), positive precedes negative (*good and bad*), and so on.

A different type of semantic constraint is the one Benor & Levy (2006) call 'Power': the element perceived as more powerful appears in slot A (where 'more powerful' is broadly intended to cover anything more important or central in a given society; not only binomials such as *prince and pauper*, where power differences are directly involved, but also *salt and pepper*, or scales of intensity, such as in *cruel and unusual*).

Another important semantic factor is the tendency for the word order of a binomial to iconically reflect the real-world order of the things or events referred to: slot A is occupied by the thing or event that comes first, chronologically and/or causally and/or along a scale (cf. *spit and polish*, *unconstitutional and severable* (only if some provisions of a law are found to be unconstitutional, the remaining provisions may be severable), *elementary and high (school)* respectively). Finally, items in slot A tend to occur more frequently in usage than items in slot B (Fenk-Oczolon 1989).

One of the most important and theoretically interesting properties of item linearization in binomials is the way the various constraints interact. Quite often, more than one constraint may be relevant, and they may favor opposite linear orders. For example, the English input binomial {*bread, butter*} is typically linearized as *bread and butter*. This order is consistent with the tendencies to have fewer syllables in slot A and to avoid stress lapses, as well as with the greater 'Power' of *bread*. At the same time, it goes against the tendency for A to have a simpler initial cluster (*br-* in slot A, only *b-* in slot B). It has been often noticed that semantic constraints outrank phonological ones when they conflict (as reported by Benor & Levy 2006 with respect to English, and Müller 1997 with respect to German). However, occasionally a phonological principle may override a semantic one, as in *trick or treat* (Cooper & Ross 1975: 73), where the iconic principle would place *treat* first ("If you don't give us a treat, we will play a trick on you"), but the preference for shorter stressed vowels places *trick* in slot A.

Another important aspect of the interaction of constraints on word order in binomials is whether they are 'strictly' ranked (i.e., if constraint X is ranked above constraint Y, Y is relevant only if orders AB and BA tie with respect to X, while an order that satisfies X can violate Y and any other constraint lower that X), or if their effect is cumulative, so that a binomial order can violate a strong constraint if it satisfies a few weak constraints (what Benor & Levy 2006 call 'gang-up effect'). In the former case, an obvious candidate to formalize their interaction would be standard Optimality Theory (Prince & Smolensky 1993/2004); in the latter case, a theory such as stochastic Optimality Theory (Boersma & Hayes 2001), Noisy Harmonic Grammar (Boersma & Pater 2016), or a 'variable rule' approach (Cedergren & Sankoff 1974) would be more appropriate. Due to the exploratory nature of the present work and space limitations, in section 7 we will only present a quantitative analysis of the strength of each constraint we propose, without trying to further explore the way they interact; see Canalis & Gök (in preparation) for a more comprehensive account.

4. Previous studies on word order in Turkish binomials. In this section we provide a brief summary of the constraints previously proposed in the literature to account for word order in

Turkish binomials. This summary does not aim to be exhaustive, but only to outline the main observations and proposals.

As mentioned in section 2, Turkish binomials are usually subsumed under the category of *ikileme*. This means most works treat them together with reduplicative prefixation, *m*-reduplication and full reduplication. In fact, one of the earlier accounts, namely Ağakay (1953), differentiates between partial and full reduplication, proposing the term *ikizleme* 'twinning' for the latter and maintaining the widely used *ikileme* for the former, and in another work (Ağakay 1954) he differentiates between reduplications and binomials labelling the latter *koşma* 'running', but his proposal to terminologically distinguish them from the other reduplicative constructions did not gain wide acceptance. Ağakay proposes some semantic and phonological criteria governing word order in Turkish binomials, which for the most part are similar to common cross-linguistic preferences mentioned in section 3. Among the semantic criteria, his proposals are versions of the iconic order (the word order of the binomial reflects arrangement of events in time, order in a sequence, and so on) and the importance order principles; he also adds that meaningless words and words with the negative suffix *-mA* are always in slot B. As for the phonological criteria, he argues that the item with fewer syllables comes first; if the syllable count of the two items is the same, vowel-initial items come first. If both items are vowel-initial, the hierarchy is /a/ > /e/ > /ɯ, i/ > /o, ø/ > /u, y/ (where '/X/ > /Y/' means 'the item starting with vowel /X/ precedes the item starting with vowel /Y/'). Ağakay does not consider the role of consonants. Finally, and again in line with observations from other languages, he says that semantic criteria are more important; whenever they are inconclusive, phonological ones step in.

Hatiboğlu (1981) is one of the most comprehensive works on *ikileme*. She lists some phonological factors governing the shape and word order of binomials, which we report in (2):

(2) a. items in a binomial tend to have the same initial consonant and the same final consonant (the latter often because they employ the same suffix(es))
b. if the vowels /a/ or /e/ occur in the first syllable of item A, then there should be /u/ or /y/ respectively in the first syllable of item B
c. if /ɯ/ occurs in the first syllable of A, the other syllables of the binomial have /ɯ/ as well

She also proposes the following constraints on their word order:

(3) a. Turkic words precede loanwords (mostly from Arabic and Persian)
b. A has fewer syllables than B
c. if there is a word starting with a vowel, it is in slot A
d. if one of the words starts with /b/ or /p/, it is in slot B

Although she does not explicitly state it as a factor, Hatiboğlu also says that sometimes the items of a binomial are ordered chronologically. She provides a long list (1956 constructions) of *ikileme*, although many of them are full reduplications and trinomials; the criteria she proposes are not tested against her whole list, but each of them is merely followed by a comparatively short list of binomials that fit it.

Tuna (1982-1983) is another detailed account of *ikileme*, and to the best of our knowledge, it is also the only quantitative study of them so far. He assembled a corpus of 2022 binomials (although most of them are no longer used in contemporary Turkish, and some of them are merely repetitions of the same binomial with different inflectional endings), for the most part irreversible and in which A and B have the same number of syllables. He does not examine binomials in

which A and B have a different number of syllables, although he states, as Hatiboğlu and Ağakay do, that if an item has more syllables it should occur in slot B.

Regarding the criteria he proposes, he rejects semantic criteria altogether, based on his rather peculiar assumption that there should only be one uniform explanation for a certain linguistic phenomenon (Tuna 1982-1983: 165). He believes that phonological constraints are sufficient to account for the word order of Turkish binomials (at least the ones he considers), and proposes this ordered ranking for his rules:

(4) a. If an item ends with a vowel and the other with a consonant, their initial segments are checked. If one is a consonant and the other is a vowel, the item ending with a vowel comes first[5]

b. If both items are consonant-final or both are vowel-final, the initial segments are checked again. If one item starts with a vowel and the other starts with a consonant, the vowel-initial one comes first[6]

c. If that is not the case, this means that the initial segments are either both vowels or both consonants. In the former case, word order depends on the following vocalic hierarchy:

/e/ > /a/ > /i/ >/ ɯ/ > /ø/ > /o/ > /y/ > /u/

In the latter case, word order depends on the following consonantal hierarchy:

/ğ/[7] > /h/ > /l/ > /r/ > /n/ > /m/ > /v/ > /j/ > /k/ > /z/ > /s/ > /ʃ/ > /f/ > /ʤ/ > /d/ > /g/ > /b/ > /ʧ/ /t/ > /p/

d. If the initial segments are identical, each segment in an item is compared with the segment having the same linear position in the other item, starting from the left, until a pair of non-identical segments occurs. If one is a vowel and the other is a consonant, the item containing the vowel will be in slot A; if both are vowels, the vocalic hierarchy in (4c) applies; if both are consonants, the consonantal hierarchy in (4c) applies

Tuna (1982-1983) reports that his set of rules fails to predict the actual word order of a mere 3.46 per cent (70/2022) of his list of binomials (although, as mentioned above, this seemingly impressive result is obtained at the cost of limiting the empirical domain of his rule set to binomials in

[5] This criterion is similar to the typologically common observation that syllables with codas, especially if word-final, usually occur in slot B.

[6] This criterion is similar to the typologically common observation that simpler syllable onsets, especially if word-initial, usually occur in slot A.

[7] This consonant – traditionally called 'soft g' – may be a velar fricative or approximant for some speakers, but for many others it has no intrinsic phonetic content; a recent acoustic study (Ünal-Logacev et al. 2019) argues that it always lengthens the preceding vowel, without having a place of its own. For such speakers, it can be analyzed as an underlyingly empty consonant position (Clements & Keyser 1983: 71). This interpretation entails that it is sensible to consider /ğ/ as one of the phonemes of Turkish, and thus include it in the sonority scale, even for varieties where it never surfaces as a separate segment. As Clements & Keyser (1983), Ünal-Logacev et al. (2019) and others observe, even if /ğ/ phonetically realized as lengthening, it is not phonologically identical to a long vowel. For example, roots ending in /ğ/ select allomorphs that must follow consonants rather than vowels (for instance, the accusative form of *dağ* 'mountain' is [da:-ɯ], like in [dał-ɯ] 'branch-ACC.', and unlike in [kira-jɯ] 'rent- ACC.').

which A and B have the same number of syllables, and of including a great number of binomials no longer in use or repeating more than once essentially the same binomial).

Apart from these accounts, more recent studies either adopt the same criteria introduced by the pioneering studies mentioned above, or have a narrower empirical domain. For instance, Çoraklı (2001) lists the following constraints: items with fewer syllables come first; vowel-initial words come first; Turkic words come first; words starting with a bilabial stop come last; the typical vowels in A and B respectively are either /e/ - /y/, or /a/ - o/, or /a/ - /u/, or /a/ - /i/, or /i/ - /a/, or /i/ - /o/ or /a/ - /y/; consonants /p, f, k, s, z, m/ tend to start item B. Baturay (2010) restricts her discussion to onomatopoeic reduplications. Based on onomatopoeic binomials in which the vowel in A is [a] and the vowel in B is [u], or in which the vowel in A is [e] and the vowel in B is [y],[8] she concludes that rounded vowels (or vowels having the U element, in the Government Phonology framework she adopts) must occur in items in slot B.

Aksoy (2019) lists *ikilemeler* in various dialects of Turkish. She observes a few tendencies already reported in previous studies: vowel-initial words come first; if both words start with a vowel, the vowel hierarchy proposed by Tuna applies; if both words start with a consonant, the consonant hierarchy proposed by Tuna applies; words with fewer syllable come first.

Finally, there are stylistic studies on binomials in literary works, and descriptive studies that compare Turkish with another language, such as Suçin (2010), comparing Arabic and Turkish binomials, and Kayalar (2017) comparing English and Turkish binomials. Neither of them introduces factors that were not already proposed before.

5. Our criteria. Based on the findings reported in the previous literature and our preliminary investigation of Turkish binomials, we tentatively adopt the following set of constraints. The only semantic constraints we assume are:

(5) 'Iconic Order': slot A is occupied by the thing or event that comes first, chronologically and/or causally and/or along a scale, e.g. *er geç* 'early (obsolete) – late' / 'sooner or later'

(6) 'Prominence': slot A is occupied by the element perceived as more prominent, important, central, powerful, common, bigger in size or higher on a scale of intensity', e.g. *çayır çimen* 'meadow – grass' / 'meadows'

In a few cases item A might also be said to be semantically less marked than item B, but we found no binomial where only markedness determines its word order, thus we did not include semantic markedness constraints. The phonological constraints we adopt are:

(7) 'Syllable Count': words with fewer syllables are in slot A, e.g. *dal budak* 'branch – shoot' / 'intricate, detailed'

(8) 'No Initial Onset': vowel-initial words are in slot A, e.g. *acı tatlı* 'sweet – bitter'

(9) 'Final Coda': consonant-final words are in slot B, e.g. *börtü böcek* 'creepy – crawly'[9]

[8] Of which only two apparently exist, i.e. *kem küm* 'onomatopoeic word' 'onomatopoeic word' / 'stammering' and *ef(il) üf(ül)* 'onomatopoeic word' 'onomatopoeic word' / 'blowing with a certain energy'.

[9] A reviewer suggests that the constraints (8) and (9) might be subsumed into a more general tendency to avoid hiatuses and clusters across slot boundaries; since (8), which prefers vowel-initial words in slot A, would make them less common in slot B, and (9), which prefers consonant-final words in slot B, would make them more common in slot A, the favoured sequence would be -V#C-. However, Turkish binomials (and more generally, Turkish phonology) do not seem to eschew clusters and hiatuses across words boundaries; a consonant-final A is fine before another consonant, provided that B is consonant-final too. In fact, -C#C- binomials abound, being more than half of

(10) 'Vowel Roundness': words with a higher number of rounded vowels are in slot B, e.g. *alt-ı üst-ü* 'below – above' / 'all in all'

(11) 'Vowel Backness': words with a higher number of back vowels are in slot B, e.g. *derme çatma* 'collection – wooden construction' / makeshift'

(12) 'Vowel Height': words with a higher number of high vowels are in slot B, e.g. *hesap kitap* 'calculation – book' / 'finances'

(13) 'Labial Consonant': words starting with a labial consonant (/p, b, f, v, m/) are in slot B, e.g. *çıtı pıtı* 'meaningless word – meaningless word' / 'mignon'

(14) 'Consonant Sonority': if the two items have the same number of syllables, compare consonants (starting from the left and ignoring word-initial consonants if the other item starts with a vowel) according to the scale /ğ, h/ > liquids and nasals > /v, j/ > /z/ and voiceless fricatives > voiced stops and /ʤ/ > voiceless stops and /ʧ/;[10] the item with the most sonorant consonant is in slot A, e.g. *soy sop* 'lineage – meaningless word' / 'kinsfolk'

Finally, following Hatiboğlu (1981) and others,

(15) 'Loanword': Turkic words precede non-Turkic words. The relevance of this factor is obvious in binomials that have two synonymous words, one Turkic and one borrowed, as in e.g. *mutlu mesut* 'happy – happy'[11]

6. Our data. Save for Tuna (1982-1983), previous studies of Turkish binomials are not supported by quantitative data. To evaluate the accuracy of the constraints we propose in accounting for the word order of Turkish binomials, we compiled a list of irreversible binomials drawn from the examples mentioned in the previous literature (mainly Hatiboğlu 1981 and Tuna 1982-1983) and the native speaker knowledge of the first author. We excluded reversible binomials, as well as fully reduplicating and *m*-reduplicating *ikileme* and binomials no longer in use in contemporary Turkish. The overall number of binomials in this list is currently 351 (additional data will be included in Canalis & Gök, in preparation).

For each input binomial {X, Y}, we coded the two corresponding possible surface binomials according to whether their word order complies with the constraint, or it does not comply with the constraint, or the constraint cannot have an impact on the order of that given binomial (for instance, Syllable Count is not relevant when both items have the same number of syllables), and finally whether the surface binomial actually occurs or not.

7. Statistical analysis and discussion of the results. In this section we examine the effectiveness of our criteria in predicting word order in Turkish binomials, and briefly comment our

---

all binomials in our list. Only one -V#V- binomial, namely *elma armut* 'apple – pear' / 'fruit' occurs in our list, but the strong preference for vowel-initial words to be in slot A is sufficient to explain their rarity: for a hiatus to occur, not only B must be vowel-initial and A must be vowel-final, but A must be vowel-initial as well.

[10] This scale is similar to Tuna's (1982-1983). Save for its lesser granularity, the main difference is the lower position of /k/, which in our scale is grouped with the other voiceless stops; we believe its higher position in Tuna's scale results from his ignoring semantic criteria, which cause an item starting with /k/ to occupy slot A in a few binomials (e.g. *kör(ler) sağır(lar)* 'blind – deaf' / 'ignorant', *kel fodul* 'bald – conceited'/ 'conceited'). The position of /j/ is discussed in section 8.

[11] The adoption of this constraint is also the reason why the preference for long vowels to occur in B was not included among the phonological constraints. Turkish words containing long vowels are usually loanwords from Arabic or Persian, therefore a constraint 'Vowel Length' would have been redundant.

results. The number of binomials that satisfy the semantic, loanword and phonological constraints are reported in Table 1, 2 and 3 respectively.

Semantic constraints	Constraint is satisfied	Constraint is not satisfied	N/A
Iconic Order	19 (82.6%)	4 (17.4%)	276
Prominence	54 (78.3%)	15 (21.7%)	230

Table 1: Frequency of binomials satisfying each semantic constraint

Loanword constraint	Constraint is satisfied	Constraint is not satisfied	N/A
Loanword	25 (86.2%)	4 (13.8%)	270

Table 2: Frequency of binomials satisfying the loanword constraint

Phonological constraints	Constraint is satisfied	Constraint is not satisfied	N/A
No Initial Onset	89 (84%)	17 (16%)	245
Syllable Count	97 (86.6%)	15 (13.4%)	239
Labial Consonant	65 (80.2%)	16 (19.8%)	270
Vowel Roundness	98 (76.6%)	30 (23.4%)	223
Vowel Height	104 (64.6%)	57 (35.4%)	190
Vowel Backness	69 (55.2%)	56 (44.8%)	226
Consonant Sonority	109 (69.4%)	48 (30.6%)	194
Final Coda	30 (51.7%)	28 (48.3%)	293

Table 3: Frequency of binomials satisfying each phonological constraint

We fitted a logistic regression model to the data using the *glm*() and *step*() functions in R 4.0.5 (R Core Team 2021). A stepwise logistic regression procedure with forward selection was conducted on all predictors. The two constraints that did not improve the model were Vowel Backness and Vowel Height. In a model with the remaining constraints, all of them turn out to be statistically significant predictors of word order (16).

(16)

Predictors	Estimate	Std. Error	z value	Pr(>|z|)
No Initial Onset	2.84	0.343	8.279	< 0.001
Final Coda	1.339	0.3448	3.884	< 0.001
Syllable Count	3.06	0.3292	9.297	< 0.001
Labial Consonant	0.8385	0.2927	2.865	< 0.01
Vowel Roundness	1.955	0.2576	7.501	< 0.001
Consonant Sonority	0.9209	0.1813	5.079	< 0.001
Iconic Order	2.623	0.4863	5.395	< 0.001
Prominence	1.344	0.2947	4.561	< 0.001
Loanword	1.151	0.5616	2.050	< 0.05

No Initial Onset, Final Coda and Syllable Count are cross-linguistically well-attested phonological constraints affecting the word order of binomials, and so are Prominence and Iconic Order

among the semantic constraints. The Loanword constraint is apparently more language specific, although it could perhaps be reduced to a semantic markedness constraint – it could be argued that native words are less marked than loanwords. As for the segmental phonological constraints, that is Vowel Roundness, Consonant Sonority and Labial Consonant, we examine them in the next section.

8. Segmental phonological constraints. The constraints Vowel Roundness, Consonant Sonority and Labial Consonant raise a few questions about their motivation and/or the phonological properties underpinning them.

8.1. VOWEL ROUNDNESS. The preference for rounded vowels to occur in slot B has already been observed for other languages. However, since most analyses are based on English or other languages in which roundness is not independently contrastive (in English for example all front vowels are unrounded, and all rounded vowels are back), it is not clear whether the preference is really about vowel roundness or backness. In fact, Cooper & Ross (1975) propose that the relevant parameter in English binomials is the second formant: vowels with a higher F_2 tend to be in slot A, yielding the scale i > ɪ > ɛ > æ > a > ɔ > ʊ > u. Turkish – where backness and roundness are independently contrastive – offers a more favorable ground to understand which property is pertinent in binomials. The data from Turkish show that the Roundness and Backness constraints have different weights, which is more consistent with them reflecting two separate properties (in fact, Backness is hardly relevant in Turkish binomials). Furthermore, it also implies that segmental phonological constraints on binomial word order are best expressed in terms of phonological features rather than raw acoustic phonetic properties. As for the reason why rounded vowels are preferred in slot B, we believe it is an instance of the general trend to place there more marked, more complex structures. Rounding is standardly assumed to be marked (see e.g. Clements 2009: 36); this fact is directly captured in features theories that adopt unary features, in which rounded vowels have a feature ([Round], [Labial] or an equivalent label) that is absent is corresponding unrounded vowels, making the former inherently more complex than the latter. In this regard, see Kabak (2007) on the motivation for a privative feature [Labial] in Turkish, mainly based on evidence from hiatus resolution processes, as well as Baturay's (2010) use of Government Phonology's U element.

8.2. LABIAL CONSONANT. This constraint appears to be rather language-specific, as word order in binomials is usually not determined by consonantal place (Cooper & Ross 1975: 72). We tentatively propose that the preference to place labial consonants in slot B in Turkish could be related to the preference for rounded vowels to occur there too, especially if the same feature is assumed ([Labial], U or any comparable formulation) to be present in both rounded vowels and labial consonants (see also Baturay 2010).

8.3. CONSONANT SONORITY. The hierarchy in (14) (reproduced in (17) for ease of consultation) raises two questions. First, which definition of sonority would be most compatible with it? Second, why is /j/ – supposedly a glide, and therefore routinely assumed to belong to the most sonorous class of consonants – lower than liquids and nasals, and on a par with /v/?

(17) /ğ, h/ > liquids and nasals > /v, j/ > /z/ and voiceless fricatives > voiced stops and /ʤ/ > voiceless stops and /ʧ/

With respect to the first question, there is no dearth of definitions of sonority and the sonority scale (see Parker 2011 and 2012 for recent summaries). Two frequently proposed definitions of its phonetic correlates are respectively based on openness of the vocal tract or (supralaryngeal)

aperture (more sonorous sounds have a more open vocal tract) and loudness (more sonorous sounds are louder, other things equal) (Parker 2011: 1178). The scale in (14) does not correlate very tightly with loudness, even leaving aside /j/. First, sibilant fricatives are louder than non-sibilant ones, yet /v/ - which is not a sibilant – is higher on that scale than any other fricative. Second, /h/ – a voiceless segment – is less loud than many other segments there. On the other hand, the criterion of openness provides a better fit; more specifically, we adopt Hume & Odden's (1996) related concept of impedance, defined as "[t]he resistance offered by a sound to the flow of air through the vocal tract above the glottis" (Hume & Odden 1996: 358), which yields the scale laryngeal > glide > liquid > nasal > fricative > stop (Hume & Odden 1996: 359). This scale nicely matches the one in (17), save apparently for the position of /j/.

This brings us to the second question. The placement of /j/ in that position on the scale appears to be empirically justified. All other things equal, the item with /j/ (spelled <*y*>) is in slot A when the other item has an obstruent (18a), but in slot B when the other item has a sonorant consonant (18b).

(18)	a.	*soy sop*	'lineage – meaningless word' / 'kinsfolk'
		yayma yazma	'dissemination – writing' / 'xx'
	b.	*kırpık kıytık*	'eyelash – meaningless word' / 'scalloped'
		helecan heyecan	'excitement – excitement'

We think the solution lies in the correct characterization of the contrastive features of Turkish /j/. Despite its phonetic realization usually being a semivowel, it can be argued that it is unspecified for the feature [sonorant]; it is never contrastive for that feature alone, and it never forms phonological classes with sonorant consonants. Several pieces of evidence support this claim: 1) Turkish allows several Sonorant–Obstruent clusters word–finally; however, compared to liquids and nasals, the distribution of /j/ is much more restricted. 2) [æ] is an allophone of /e/ before tauto–syllabic liquids and nasals, but not before /j/. 3) The distribution of the allophones of /j/ is similar to that of the allophones of /v/. 4) A preliminary phonetic survey suggests that Turkish /j/ (as well as the other [+continuant] consonants) may display friction and partial devoicing, at least utterance–finally (see Canalis et al. (to be published) for a discussion of these arguments).

The underspecification of Turkish /j/ and /v/ with respect the feature [sonorant] is consistent with their role in binomial word order as well. If they are unspecified for that feature, their intermediate position along the sonority scale between the [+sonorant] /ğ/, /h/[12], liquids and nasals and [–sonorant] obstruents directly follows, fully in line with the hierarchy in (17).

9. Conclusions and directions for future research. Our exploratory quantitative investigation of Turkish irreversible binomials suggests that the position of an item within a given binomial is affected by, at least, 1) it being a loanword, 2) the presence of an onset in the initial syllable, 3) the presence of a coda in the final syllable, 4) the presence of an initial labial consonant, 5) the presence of rounded vowels 6) a higher syllable count, 7) the lower sonority of its consonants, 8) the lower prominence of its meaning, and finally 9) its real-world referent being chronologically or logically later. All these factors concur to place an item in slot B. We also found that the evidence provided by the segmental constraints may shed light on some aspects of the phonological system of Turkish, i.e. the markedness of rounded vowels and labial consonants, the properties of the sonority scale in Turkish and the featural specification of Turkish /j/.

[12] Under the definition of sonority as impedance we adopt, glottal consonants are [+sonorant], since the airflow is not impeded above the glottis during their production.

Our constraint list is not meant to be exhaustive, our dataset could be expanded further, and we only briefly touched upon the nature of the interaction among constraints. We leave all these goals and topics for future research.

References

Aksoy, Özlem Merve. 2019. *Türkiye Türkçesi Ağızlarında İkilemeler*. Kırıkkale: Kırıkkale University MA thesis.

Ağakay, Mehmet Ali. 1953. İkizlemeler üzerine. *Türk Dili* 2 (16). 189-191.

Ağakay, Mehmet Ali. 1954. Türkçede kelime koşmaları. *TDAY – Belleten*. 97-104.

Baturay, Semra. 2010. Apophony in Turkish. *Turkish Studies* 5(4). 201-220. http://dx.doi.org/10.7827/TurkishStudies.1787.

Birdsong, David. 1979. *Psycholinguistic Perspectives on the Phonology of Frozen Word Order*. Cambridge, Mass.: Harvard University Dissertation.

Benor, Sarah B. & Roger Levy. 2006. The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82(2). 233-278. https://doi.org/10.1353/lan.2006.0077.

Boersma Paul & Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32. 45-86. https://doi.org/10.1162/002438901554586.

Boersma, Paul and Joe Pater. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In John J. McCarthy & Joe Pater (eds), *Harmonic Grammar and Harmonic Serialism*. Sheffield, Equinox. 389-434.

Canalis, Stefano, Semra Özdemir, Utku Türk & Ümit Tunçer. To be published. The phonological nature of the Turkish front glide. In *the 20th International Conference on Turkish Linguistics. The Eskişehir Meeting*. Wien: Otto Harrassowitz.

Canalis, Stefano & Serra Gök. In preparation. Frozen binomials in Turkish. Manuscript.

Cedergren, Henrietta & David Sankoff. 1974. Variable rules: Performance as a statistical reflection of competence. *Language* 50. 333-355. https://doi.org/10.2307/412441.

Clements, George N. & Samuel Jay Keyser 1983. *CV Phonology: A Generative Theory of the Syllable*. Cambridge, MA: MIT Press.

Clements, George N. 2009. The role of features in phonological inventories. In Eric Raimy & Charles E. Cairns (eds), *Contemporary Views on Architecture and Representations in Phonology*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262182706.003.0002, accessed 16 July 2023.

Cooper, William & John Ross. 1975. World order. *Chicago Linguistic Society* 11(2). 63-111.

Çoraklı, Şahbender. 2001. Türkçenin yaratma gücü ikilemeler I. *Atatürk Üniversitesi Türkiyat Araştırmaları Enstitüsü Dergisi* 17. 53-60.

Green, Viola & David Birdsong. 2018. Intuitions for phonological constraints in binomials: A psycholinguistic investigation. *Language Sciences* 66. 116-134. https://doi.org/10.1016/j.langsci.2018.02.004.

Fenk-Oczlon, Gertraud. 1989. Word frequency and word order in freezes. *Linguistics* 27. 517-556.

Gordon, Matthew K. 2006. *Phonological Typology*. Oxford: Oxford University Press.

Hatiboğlu, Vecihe. 1981. *Türk Dilinde İkileme*. Ankara: TDK Yayınları.

Hume, Elizabeth & David Odden. 1996. Reconsidering [Consonantal]. *Phonology* 13(3). 345-376. http://www.jstor.org/stable/4615491.

Jacobsen Jr, William H. 1982. Basque copulative compounds: A problem in irreversible binomials. *Berkeley Linguistics Society (BLS)* 8. 384-397. https://doi.org/10.3765/bls.v8i0.2025.

Jespersen, Otto. 1905. *Growth and Structure of the English Language*. Leipzig: B.G. Teubner.
Kabak, Barış. 2007. Hiatus resolution in Turkish: An underspecification account. *Lingua* 117(8). 1378-1411. https://doi.org/10.1016/j.lingua.2006.04.010.
Kayalar, Fethi. 2017. Comparison of structural features of binomials in Turkish language and English language. *Turkish Studies* 12. 311-324. http://dx.doi.org/10.7827/TurkishStudies.12473.
Kaye, Alan S. 2009. Cultural ingredients in Arabic lexical pairs (binomials). *Word* 60. 65-78. https://doi.org/10.1080/00437956.2009.11432593.
Lambrecht, Knud. 1984. Formulaicity, frame semantics, and pragmatics in German binomial expressions. *Language* 60(4). 753-796. https://doi.org/10.2307/413798.
Malkiel, Yakov. 1959. Studies in irreversible binomials. *Lingua* 8.113-160.
Masini, Francesca. 2006. Binomial constructions: Inheritance, specification and subregularities. *Lingue e Linguaggio* 2. 207-232. https://doi.org/10.1418/23144.
Mollin, Sandra. 2014. *The (ir)reversibility of English binomials*. Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/scl.64.
Müller, Gereon. 1997. Beschrankungen fur Binomialbildungen im Deutschen. *Zeitschrift fur Sprachwissenschaft* 16(1/2). 5-51.
Oakeshott-Taylor, John. 1984. Phonetic factors in word order. *Phonetica* 41. 226-237. https://doi.org/10.1159/000261729.
Parker, Steve. 2011. Sonority. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice, *The Blackwell Companion to Phonology. Vol. 2*. 1160-1184. Malden, MA: Wiley-Blackwell.
Parker, Steve (ed.). 2012. *The Sonority Controversy*. Berlin & Boston: De Gruyter Mouton.
Prince, Alan and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell (2004). Technical Report, Rutgers University Center for Cognitive Science and Computer Science Department, University of Colorado at Boulder (1993).
R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Suçin, Mehmet Hakkı. 2010. Turkish and Arabic reduplications in contrast. *Australian Journal of Linguistics* 30(2). 209-226. https://doi.org/10.1080/07268601003678627.
Ünal-Logacev, Özlem, Marzena Żygis, & Susanne Fuchs. 2019. Phonetics and phonology of soft 'g' in Turkish. *Journal of the International Phonetic Association* 49(2): 183-206. https://doi.org/10.1017/S0025100317000317.
Tuna, Osman Nedim. 1982-1983. Türkçenin sayıca eş heceli ikilemelerinde sıralama kuralları ve tabiî bir ünsüz dizisi. *Türk Dili Araştırmaları Yıllığı - Belleten* 30-31. 163-228.
Wright, Saundra, Jennifer Hay & Tessa Bent. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 44. 531-561. https://doi.org/10.1515/ling.2005.43.3.531.