

***Aradınız kişi (şu anda) ulaşamıyor!** The Turkish vowel system, (the so-called) 'Yumuşak g' (ğ), and Turkish phonology: On a missed opportunity**

Nicolas Royer-Artuso

University of Quebec at Chicoutimi (UQAC), Canada

French Institute for Anatolian Studies (IFEA), Istanbul, Turkey

Centre for Research on Musical Traditions, Antonine University, Beirut, Lebanon

Abstract

In this paper, some core premises that are held about Turkish Phonology are put into question, both theoretically and empirically. Some modifications to the Turkish phonological inventory and to the language's phonotactic constraints are then proposed. It is shown how modifying the phonological inventory and modifying phonotactic statements about the language gives a more realistic perspective on the empirical data. In the conclusion, some new avenues of research are finally proposed.

Keywords: phonology; morphology; phonology-morphology interface; Turkish; language change

1. General introduction

In this paper, I propose some modifications to some core premises that are held about Turkish phonology:

1. A modification of the inventory of phonemes:
 - a. I will argue that the so-called 'Yumuşak g' (ğ) is not a phoneme of the language anymore, and that the surface structures, in which it never appears, should be counted

* *Aradınız kişi şu anda ulaşamıyor!* is the automatic answer we get from some telephone companies in Turkey when the cellphone of the person we call is busy or off. The literal translation is 'The person you are calling is not joinable at the moment'. However, *aramak* is a verb that also means 'to search for' and therefore, *Aradınız kişi şu anda ulaşamıyor!* can also mean 'The person you are *searching for* is not joinable at the moment'. Which, as I will try to show in the present paper, is exactly the case for Turkish so-called *yumuşak g* (ğ). Note that the 'literal/orthographic' form of *Aradınız* is *Aradığınız*, a form supposedly containing *yumuşak g*, the topic this paper revolves around.

The author wants to thank the CRSH for funding, and Luc Baronian and the reviewers for helpful comments.

as phonemically valid, i.e. ğ is *phonetically* absent, and I argue that it is *phonemically* absent as well;

- b. I will also argue that long vowels are now part of the language's phonemic system;
2. A modification of the phonotactic constraints of the language, namely, on the constraint on clusters of vowels, which I argue does not exist in the present state of the language.

In order to do that, I will connect the problems of Turkish ğ with other well-known problems in Turkish phonology, this way showing a) the premises' co-dependencies and b) how the proposed modification permits rethinking other problems that do not seem at first sight to be connected. Section 5 offers some notes on the theory of phonology that connect the discussion of the present phonological problem (i.e. *yumuşak g* 'soft g' (ğ)) with premises that are generally held, and that probably have to be reconsidered.

The aim of this paper is to come up with a better analysis of the empirical data. At the same time, all the discussions that are found below touch on methodological foundations of the science of Phonology; on some of the premises that are held; and on the way these premises constrain the way we develop our phonological arguments.

2. The problem with Turkish *yumuşak g* 'soft g' (ğ)

In reference grammars of Turkish, we find statements like the following:

- i. The so-called 'soft g' lacks a corresponding 'consonantal' sound in standard Turkish [...] When it is in word-final or syllable-final position, it lengthens a preceding back vowel [...] Between identical back vowels it is inaudible [...] (Göksel and Kerslake 2005: 7).
- ii. Long vowels occur in words borrowed from Persian and Arabic (Ibid.: 11). There are no vowel sequences in Turkish, except in loan words (Ibid.: 12).

(i) and (ii) contradict each other. What is described in (i) often corresponds phonetically to a long vowel and *do occur* in 'native' words, e.g. *da* 'mountain'; *si* 'shallow' (etymologically/orthographically *dağ* and *sığ*, respectively).

These are some of the 'effects' that are attributed to *yumuşak g* 'soft g' (ğ).

The so-called 'soft g' lacks a corresponding 'consonantal' sound in standard Turkish, although it is pronounced as a voiced velar fricative in some dialects. It behaves like a consonant when a suffix follows it [...], and is either inaudible as a consonant or may be pronounced as a palatal glide in the environment of front vowels and as a bilabial glide in the environment of rounded vowels. In particular:

- (i) When it is in word-final or syllable-final position, it lengthens a preceding back vowel (*dağdan* [da:dan] 'from the mountain' and *sığ* [sɪ:] 'shallow'), but may be pronounced as a palatal glide when following a front vowel (*eglen-* [ejlæn] 'have fun').
- (ii) Between identical back vowels it is inaudible (*sığınak* [sɪ:nak^h] 'shelter', *ugur* [u:ɣ] 'good luck').
- (iii) Between identical front vowels it is either inaudible (*sevdiğim* [sevdi:m] 'that I love') or sounds like a palatal glide (*düğün* [dүйүн] 'wedding').
- (iv) Between rounded vowels it is mostly inaudible but can also be pronounced as a bilabial glide *soguk* ([souk^h] or [soʊuk^h] 'cold').
- (v) Between a rounded vowel and an unrounded vowel it is mostly inaudible but can also be pronounced as a bilabial glide (*sogan* [soan] or [sowan] 'onion').

- (vi) 'a+ğ+i' sequences may either sound like a sequence of /a/ followed by /ɪ/ or like a sequence of two /a/ vowels (*agır* [aɪɾ] or [a:ɾ] 'heavy').
- (vii) 'ı+ğ+a' sequences, on the other hand, are pronounced as sequences of /a/ followed by /ɪ/ (*sığan* [sɪan] 'one' which fits').
- (viii) When 'ğ' occurs between an 'e' and an 'i' it is either inaudible or pronounced as a palatal glide [j], hence words with the sequences 'e+ğ+i' and 'i+ğ+e' can sound like words written with a 'y', as in *değil* [dejil] 'not' and *diğer* [dijæɾ] 'other' (the former being similar to words *written* with a palatal glide, e.g. *meyil* 'slope'). 'e+ğ+i' sequences may also sound like a sequence of two /i/ vowels, hence *değil* is often pronounced [di:l] in colloquial speech.

Göksel and Kerslake (2005: 7)

While *ğ* has received some attention in the literature on the phonology of this language, I argue that some of its properties have not been properly connected to other phenomena in the language. This constitutes one of the missed opportunities to understand Turkish's vowel system and the language's active phonology: when we take into account the *synchronic* properties of *ğ* (i.e. when we do not map diachrony onto synchrony), some other problematic facts of the language receive straightforward solutions (but we also need to revise many other things accordingly).

A general note is needed before starting. Turkish phonology is for the major part very simple, one could say 'ideal' (that is: before looking in details at the data, and especially at the 'exceptional' cases). The analyses found in the literature about Turkish phonology are therefore rather traditional and straightforward. The phonological representations are phonemic and linear¹, and the majority of processes very easy to formalize using phoneme replacement rules given a certain phonological context (i.e. $x \rightarrow y/z$).

3. Turkish Vocalic System

This is the vowel system ('repertoire' of distinctive vocalic segments) generally given for Turkish (1):

(1)	Unrounded		Rounded	
	Open	Close	Open	Close
Back	a	ɪ	o	u
Front	e	i	ö	ü

In reference grammars², not far from this phonological 'fact', the reader will find statements such as the following:

¹ Phonemes could of course be replaced by features and/or linear representations by non-linear ones. The important point being that this would not affect immensely what is said about Turkish phonology.

² I will use in the present paper three Turkish grammars interchangeably for the main part of my demonstration, i.e. Kornfilt (1997), Lewis (2000) and Göksel and Kerslake (2005). The reason is that what these grammars have to say about the problems I discuss is to a large extent identical (minus very little details that do not affect the present discussion), and that at the same time, the way they present the data is the background on which phonologists more or less agree. To the best of my knowledge, no one has yet presented an analysis like mine.

Turkish vowels are normally short but may be long in three situations [...] The difference between short and long vowels is of quantity not quality: the positions of the speech organs is the same; the change is in the length of time during which breath flows.

Lewis (2000: 21-22)

Three situations in which long vowels occur are:

- a. Foreign borrowings
- b. Any vowel followed by ğ + consonant
- c. When it is desired to emphasize a word, one vowel may be dwelled on, i.e. lengthened.

Ibid. (2000: 30-31)

Native Turkish vowels are short *phonemically* [emphasis mine]. They can be lengthened due to processes of compensatory lengthening [...] Borrowed words can have vowels with phonemic length, however (Kornfilt 1997: 489) [...] Turkish has essentially only short vowels. However, as stated in that subsection, some loanwords do have long vowels, and the native phonology itself can give rise to long vowels via processes of compensatory lengthening. As a result of these facts, vowel length can be distinctive, at least auditorily: dağ 'mountain' [d a:] da 'also, too' [d a] [...] saat 'hour watch' [s a a t] or [s a: t] sat! 'sell' [s a t].

Ibid. (2000: 501)

In these quotes, we clearly see the connection between loanwords and long vowels: *phonemic* long vowels, it is said, appeared in Turkish only as a result of borrowing. For native words, the surface long vowels can only be attributed to a phonological process. Compensatory lengthening is often the given explanation: phonemically, these vowels are short, and point to a "(semi-abstract) segment (Kornfilt 1997: 488)" lengthening them at the output. The reason we have to postulate ğ is that:

there is a phonemic consonantal segment which remains unpronounced in the standard language. Because it does have certain phonological effects, [...] it must be recognized in the phonemic inventory of the phonemic consonantal segments. In some dialects, this segment is pronounced as a voiced velar fricative [ɣ]. Let us mention its main phonetic effect: when it is in syllable final position and cannot be resyllabified with a following vowel, it triggers compensatory lengthening of the preceding vowel: çağdaş 'contemporary' [tʃ a: d a ʃ].

Ibid.

Below, I will show how another connection is to be made with the current-mainstream description of Turkish phonotactics, this time concerning its syllabic structure.

4. Reasons for the current analysis of Turkish vowel system

No reasons other than

1. the Phonemic Principle;
2. what we could call the 'Contrastive Principle' and the 'Complementary Principle'; and
3. Etymology,

are responsible for the current synchronic analyses of Turkish ğ.

1) By the Phonemic Principle here, I mean the common technique of finding the phonemes of a language by finding minimal pairs of morphemes/words that only differ by one sound. If

we can find such morphemes-words, we can assume that underlyingly, the two sounds exist in the system.

In the case of *ğ*, finding minimal pairs consisting of one word with and one word without a long vowel won't give much results for so-called 'native words' (but some pairs can actually be found, e.g. *aç* 'hungry' and *ağaç* (/āç/) 'tree' are in opposition). With so-called 'borrowed words', it will be easier, but the strategy will be to make a distinction between native and non-native words (a strategy that for Turkish occupies an important remedy to many of the problems posed by so-called non-native words, e.g. Vowel Harmony and Lexical Stress). The rare occasions when a long and short vowel make a difference in words will be sent, circularly, to deep representations: basically attributed to *ğ* (e.g. *ağaç* /āç/ has a *ğ* phonemically).

2) By the 'Contrastive Principle' and the 'Complementary Principle' I mean the method that consists in comparing words with and without their affixes, and from there, to come up with the deep *invariant* representations of these morphemes plus the rules that permit to explain the found ('surface') allomorphy. For the present discussion: cases where a) we assume *ğ* only because we compare words with and without their affixes (in Turkish, for the overwhelming majority, suffixes); b) find which allomorph is the 'basic' morpheme; and c) posit *ğ* in the input of a rule.

One important topic in Turkish Phonology is the case of the so-called 'k-Ø alternation' (where Ø=*ğ*). Classical examples include:

- (2) köpek 'dog' + acc. '-i' → köpeğ-i 'dog /köpei/;

and for complex words involving more than one affix:

- (3) yap 'to do' + -acak 'fut.' + -m '1st pers. → yapacağım /yapacağım/

(where *ı* is said to be the result of an epenthetic rule).

By postulating that a) the input forms contain a *k* and b) that these forms are invariable phonemically, we need to postulate the (morpho)phonological rule *k* → *ğ*, a rule turning *k* into *ğ* and then turning underlying *ğ*'s into surface Øs, thus permitting to keep *ğ*'s in our phonemic system. And thus preserving symmetry between etymological and modern representations.

3) The role of etymology for this analysis is primordial. *ğ* is first of all a phoneme that can be found in earlier states of the language as well as in some modern dialects. It is thus convenient to attribute a phonemic role to *ğ* – even if its phonetic properties do not militate for it.

The etymological bias for *ğ* can also be thought in parallel with the way borrowed words are treated: nowhere does an analysis of borrowed words with long vowels exist in which these long vowel are treated as being composed of two identical vowels separated by *ğ* (an analysis that is preferred for native words);

4) All of the above (1-3) are also the result of a fundamental decision about Turkish phonology/phonotactics, a decision that has repercussions in the treatment we give of many related and unrelated phonological facts. This decision concerns the syllabic structure of Turkish:

The most common combinations of consonants (C) and vowels (V) are VC (at 'horse', ol- 'be', in- 'descend', iş 'work', üç 'three') and CVC sequences (gel- 'come', bak- 'look', güz 'autumn', göl 'lake'). There are also CV

sequences such as *bu* 'this', *şu* 'that', *su* 'water', *ne* 'what', *de-*, 'say', but these are fewer in number and those which are nominals usually require an additional consonant ('n' or 'y') when they combine with suffixes (6.1.3). Other types are VCV sequences such as *ara-* 'look for', *ile* 'with', *öte* 'far side', and VCC and CVCC sequences such as *ört-* 'cover', *sert* 'hard', *genç* 'young'.

Göksel and Kerslake (2005: 12)

What this means, in terms of the present discussion, is that no cluster of vowels can be found in the same syllable. This applies much more in the discussions about suffixes. In monomorphemic words, it's not as necessary to mention this, except etymologically, e.g.:

In Turkish roots are predominantly monosyllabic, i.e. they contain a single vowel. There are no vowel sequences in Turkish, except in loan words.

Göksel and Kerslake (2005: 12)

For morphological operations involving native morphemes, it is argued, there must be (logically) a consonantal buffer to avoid two vowels to appear next to each other (as in the case of the 'k-Ø alternation'). If not, how do we account for the found ones without putting into question many of the premises we have postulated?

5. Notes on phonological theory and some of its premises

In the analysis of a language, the work of the phonologist can be roughly divided into two major tasks:

1. giving an account of the system of phonemes of the given language and the different manners these phonemes are organized at the syntagmatic level; and
2. giving an account of the rules or repair mechanisms that the language possesses.

(2) necessarily presupposes (1), that is to say: we cannot account for phonological processes, may they be morpheme-internal (more or less: allophonic), or morphologically – or syntactically – activated (more or less: allomorphic) unless we know how the inputs to these processes are represented.

In other words, this means that:

1. giving an account of the system of phonemes of a given language and the different manners these phonemes are organized at the syntagmatic level

CONDITIONS

2. giving an account of the rules or repair mechanisms that a language possesses.

A given analysis of the system of phonemes has some necessary logical implications, and this, circularly, only *because* we have analysed it this way. That also means that the problematic residues, e.g. exceptions and hard to analyze data, are also bound by the whole analysis, and this makes it extremely difficult to get out of the logical constraints that we have set up, so to speak, in advance: something we said in the premises will have dramatic repercussions on the following argumentation.

While this observation is pretty banal and what is described arguably necessary as a scientific method, some other more hidden methodological premises also exist that reinforce the way we analyze data and, coupled with what was presented above, carry many other logical implications (presented below with no specific hierarchical importance):

1) The first logical implication has to do with the distinction between synchronic phonological rules and the result of these processes, that is: an emphasis on etymological representations even when it is clear that these representations have changed. Etymological *residues* pose problems in the sense that they disrupt the behaviour of the postulated rules at the level of the entire system, i.e. the residues tell a story about the past, a story that does not hold for the present. These residues thus behave in ways that are unexplainable if we do not assume etymological representations – when of course we work with the accepted inventory of phonemes and phonological rules. And that is precisely the reason why the etymological representations need to be assumed. But in doing so, we inevitably distort some of the facts;

2) The second logical implication has to do with the distinction between native *versus* borrowed words. This emphasis amounts in many cases to a (light) form of essentialism. The question of the division between native and non-native words is tied closely to the question of etymological forms. When we accept that 'borrowed' words can be counted as normal 'native' words for the native speakers, our analysis of the phonology of the language can then be changed accordingly, and some of the 'mysteries' become, so to speak, less mysterious;

3) More importantly – since this somehow encompasses everything else – I will try to show how an emphasis on the syntagmatic dimension over the paradigmatic one ties all these problems together.

While we need paradigms to discover the phonemes of a language – to get complementary distributions – the goal is mainly to give a syntagmatic account of the series of phonemes that constitute morphemic-lexical inputs (i.e. deep representations). From there, we go on and discover the syntagmatic constraints on the alignment of phones and phonemes and move on to the syntagmatic constraints on the phonological properties of morphological parts. Eventually, we move on to the syntagmatic properties of words at the syntactic level.

At the same time in the mainstream analyses a methodological emphasis is given to the search of *invariant* deep representations that are then affected by the phonological form of surrounding morphemes when we put them together syntagmatically. Much of the phonological rules of a language are proposed to account for the divergence between a proposed deep representation and its surface outputs in function of phonological contexts – the same way phonemes and allophones are distinguished. However, when we look at paradigms, we observe behaviours that show a very different organisation of the system of morphemic/lexical entries in terms of the phonological properties of their representations. This points to the need for a different analysis of the system of representations.

Exemplarist models³, for example, with their very different conclusions about phonological representations point to some very different analyses of the data. While these types of models lack in some of their explanations, they still possess the capability of taking into account the paradigmatic dimension, and this, because they take surface outputs seriously, both at the level

³ See e.g. Bybee (2001); Pierrehumbert (2001) and (2010), amongst many others.

of a) phonetics and the level of b) full words. In that, they join the group of morphologists that emphasize the need for analyses based on full whole-words⁴ – may they be morphologically simple or complex – constructions⁵, and/or the group of morphologists working with analogical modelling⁶ (because analogy needs whole forms to work with).

Finally, I want to add that the question of the level of abstraction we assume and/or accept in our models and theories is clearly related to what I said in the precedent discussion. However, it is not as central as it could seem: my quarrel is not with abstraction per se. It is rather with analyses that propose more and more abstraction as the phonological system changes and is restructured accordingly, abstractions needed only because the system has changed. What I propose is compatible with abstract analyses if the changes I propose are taken into account (because the current analysis of Turkish is clearly not in touch with the actual facts of the language given in the data). It is also compatible with models that are word-based as well as morpheme-based (but only if we accept more suppletion in the latter case). In that sense, what I propose is not related to any specific theory or model, but it has implications for what theories and models will do with the Turkish data and their analyses.

6. Changing some of our premises: a solution to the related problems clustered around ğ

I want to propose to rectify some premises about Turkish phonology and this, in aligning our phonological analyses with what is observed in the 'surface' data.

(1-3) are the proposed modifications:

1. there is no ğ in Modern Standard Turkish, neither phonetically (which is an accepted fact⁷) nor phonemically;
2. long vowels are now part of the phonemic inventory of Turkish; and
3. the constraint on clusters of vowels does not hold synchronically.

The reason we find in the data

- a. synchronic effects of ğ;
- b. only rarely long vowels in Turkish native words; and
- c. only rarely clusters of vowels,

is not because of phonological constraints present in Modern Turkish. Rather, it is because of residues from the past, when the language still had ğ and these constraints⁸.

⁴ See e.g. Aronoff (1976), (1994) and (2007); Anderson (1992); Bochner (1993); Ford and Singh (1991) and Ford et al. (1997); Stump (2001); Blevins (2006) and (2016), amongst many others.

⁵ See Booij (2010).

⁶ See e.g. Skousen (1992); and the collection in Skousen et al. (2002).

⁷ See Kılıç and Erdem (2008) (and the references therein), which is a recent very detailed discussion of ğ at the phonetic level. The argumentation is basically an attempt to save ğ phonemically.

⁸ Similar conclusions about synchronic phonology and historical residues can be found in Blevins (2004).

Once we abandon or at least question some of the main mainstream principles of the phonological method⁹ (presented in the precedent section) – as many frameworks in phonology and/or morphology tend more and more to do – we get a different picture of Turkish phonology and its processes. But for this, we need to reconsider some of our assumptions about Turkish's vowel system and phonotactics, as well as its morphemic-lexical representations.

The phoneme *ğ* is posited for Modern Turkish for three main reasons:

1. to *not* have to say that there are long vowels in Turkish native morphemes/words;
2. to keep up with etymology and not have to treat lexical/morphemic representations and problematic behaviours of morphological data in ways that would 'harm' the supposed symmetry of the system; and
3. to not have to suppose (weak) *suppletion* for morphemes – which is related to (1) and (2) indirectly, but is one of the core presuppositions about morphology and its relation to phonological output forms.

The absence of long vowels in Modern Turkish is posited for three reasons:

1. mainly because of etymological considerations: it is true that historically native words do not possess them; and it is true as well that outputs of word-formation processes did not give an opportunity for the apparition of such units;
2. because of the effects of long vowels on some phonological rules in the language (e.g. Vowel Harmony. See below); and
3. because if we were to assume long vowels, we would have to modify many premises we hold about Turkish phonology, including the existence of *ğ*.

Once we accept that long vowels are now part of Turkish's inventory (1), it becomes possible to take some of the encountered surface forms (considered 'output forms' in the literature) as basic-underlying forms. This is the case for all the surface forms for which a *ğ* is traditionally posited to explain the surface long vowels in 'native' words (e.g. the native orthographic ağaç 'tree' /āç/; or orthographic dağ /dā/).

The next logical step is the following: once we have accepted long vowels in our inventory, we can safely argue that *ğ* is not anymore a phoneme in the language. Every 'effect' it is said *ğ* has on the derivations from deep to surface representations can be accommodated directly in the lexicon and the representations it contains, i.e. a) by assuming new/reinterpreted phonological representations for morphologically simple words and morphemes; b) by assuming suppletive forms for the bases that show allomorphic behaviours, if we work with morphemes; or by assuming analogical word-formation processes based on paradigmatic relations between full words in the lexicon, if we work with full words. In other words, we can change the phonological representations of the input forms – without recourse to *ğ*.

Arguing that *ğ* is not a phoneme of Turkish also permits a new perspective on other phenomena of the language, phenomena that are difficult to offer an explanation to if we stick with the traditional analysis of Turkish phonology.

⁹ See e.g. van Oostendorp (2013) for a list of mainstream phonological evidence and many of their problems.

7. Turkish 'k-Ø alternation'

Pöchtrager (2013), working inside the Government Phonology framework, shows that the traditional analysis of the 'k-Ø alternation' (Ø=ğ) as the result of a productive phonological rule/process does not hold:

- (4) köpek 'dog' + -i 'acc.' → köpeğ-(ğ)i /köpei/
 bebek 'baby' + -e 'dat.' → bebe-(ğ)e /bebee/.

It is *morphological*, however we want to analyze/formalize what is happening. As Pöchtrager writes:

This article looks at the Turkish k-Ø alternation and questions whether it can be counted as a phonological process. After a discussion of what it means to be phonological, the empirical facts are weighed against the theoretical expectations. The alternation has several quirks, and any account treating it as phonological must allow for complex machinery to deal with those, thus weakening the predictive power of the theory employed. This is true of earlier rule-based accounts, but also of accounts that have been presented within Government Phonology, the framework that also this article is couched in. After careful revision of several problems that previous accounts have run into and that seem insurmountable, we must conclude that the k-Ø alternation does not qualify as phonological.

Pöchtrager (2013: 87)

Pöchtrager works with Government Phonology, which accepts no exception to a phonological process. However, as Pöchtrager rightly underlines, the data are problematic for all the models that treat the 'k-Ø alternation' as phonological: the rule is productive but gives different types of outputs depending on the inputs.

Inkelas (2011) (see pp. 84; 86 and 97) reached very similar conclusions regarding this Turkish 'phonological rule' in the sense that no 'normal' rule works smoothly when we look at the data. Her solution is to treat the problematic data with *co-phonologies*, thus allowing the treatment of different inputs with different phonologies. However, as Kabak and Vogel (2011) clearly show, co-phonology is a problematic device for Turkish, as analyses involving co-phonologies cannot explain why words often end-up participating in different co-phonologies at the same time.

If we further modify our statements about the Turkish system of vowels as well as our statements about its phonology according to what I proposed, there is no 'k-Ø alternation' problem remaining – which does remain for Pöchtrager, since he is concerned with the phonological problem but not with the way it should be treated at the morphological level. Building on Pöchtrager's analysis, we can safely say that the 'k-Ø alternation' (Ø=ğ) is the remnant of previous states of the language that have been somehow reanalyzed as suppletive – and not as a synchronic process of erasing a k phonologically and replacing it with a buffer consonant, ğ. There is now, *at least*, two (weak) suppletive forms for the root (e.g. köpek – köpe) that are chosen when affixation of specific suffixes is involved.

However, the most straightforward analysis is to posit full forms, one without and one with the suffix, and take the relation between the two and the 'rule' that can be created by comparing them as being – in some sense of the term – *analogical*, i.e. to create a new affixed form, we compare the two full words that we have stored, and analogically create a new form which lacks

a k – basically, through a kind of backformation process – a process that needs full words to take place ($Xk \leftrightarrow Xi$, e.g. *köpek* \leftrightarrow *köpei*¹⁰).

For this, another premise about Turkish phonology that should be corrected is the supposed constraint on series of vowels that are currently analyzed as being separated by *ğ*. Opening the possibilities for representations containing un-buffered series of vowels allows for more realistic accounts of morphological operations (e.g. *köpek* 'dog' + -i 'acc.' \rightarrow *köpe-i*; *kopek* + e 'dat.' \rightarrow *köpe-e*). At the same time, by postulating long vowels for native words, both morphologically simple and complex, we can thus account for long vowels that are sometimes 'created' in the process (*köpē* 'dog_[gen.]' is in free variation with *köpee*)¹¹.

All this becomes even clearer when we add a recent problem in Turkish phonology-morphology interface, i.e. the treatment of the Turkish diminutive suffix *-cik*.

8. Turkish diminutive *-cik*

This is the kind of data I am referring to when talking about the *-cik* suffix:

(5)

- a. *deve* 'giant' + *-cik* 'dim.' \rightarrow *devecik* = 'small giant' (a kind of pear that is sold in Turkey)
ambülans 'ambulance' + *-cik* \rightarrow *ambülanscık* 'small ambulance'
- b. *Nikola* 'proper name' + *-cik* \rightarrow *Nikolacık* 'small Nicolas'
Nikolacık + -m 'poss.' \rightarrow *Nikolacım* 'my small Nicolas' (orthographic *Nikolacığım*).
 Note: According to the mainstream analysis, $k \rightarrow \check{g}/_ - V(C)_{[suffix]}$
- c. *köpek* 'dog' + *-cik* \rightarrow *köpecik* 'small dog'
bebek 'baby' + *-cik* \rightarrow *bebecik* 'small baby'

Much of the analytical problems concerning *-cik* rest on an emphasis on the syntagmatic dimension that makes us blind to paradigmatic effects. One of these effects is to be found in other so-called phonological processes that affect 'underlying' forms and create allomorphs. I will argue that the solution to the *-cik* suffix is to be found paradigmatically when we look at what happens elsewhere in the system.

The last two examples in (5) are those that have occupied linguists, e.g. Inkelas (see Inkelas (2011) and (2014)). It is very difficult to offer a traditional account of what is happening there, i.e. an analysis based on invariant morphemes and phonological rules to explain the

¹⁰ $X \leftrightarrow X$ is a formalism that can be found in e.g. Ford and Singh (1991) (see also Ford et al. (1997)) and Booij (2010). It corresponds to 'Word Formation Strategies' in the former model and 'Schemas' in the latter. The variables X are occupied by related full words, may they be morphologically simple or complex. The word-formation process is based on the difference that is found between the left and right side of the arrow, and helps applying the same strategy/schema to new words that are to be created, e.g. morphologically simple words from complex words as well as the opposite.

¹¹ In Royer-Artuso (2016) and (2017), I show how it is possible to account with morphological word-based models for different Turkish phonological facts that resist traditional analyses based on morphemes. Royer-Artuso (2016) also shows the *necessity* of positing full words as the basic units in the analysis of Turkish.

allomorphy. The problem is, again, to explain why the final *ks* of *köpek* and *bebek* disappear¹². There is no phonological context that can explain and/or predict what is occurring, because:

1. **köpekçik* and **bebekçik* are phonologically well formed. (Note that $c \rightarrow \check{c}$ in suffixes/C[voice]___, a process that is totally regular);
2. The number of syllables also does not count (i.e. it is not the case that truncation takes place because *-cik* needs a bi-syllabic 'stem' cf. *ambiülans*, *Nikola*);
3. The recourse to co-phonologies is also impossible, since these words act normally according to other phonological processes.

The conclusion is that there is no real *phonological* explanation for what is happening, at least from a morphemic point of view. We can say that there is consonant deletion, but it is not a general rule: only by listing these words in their morphologically simple as well as complex forms in the lexicon could we save the general analysis, i.e. to treat them as exceptions, therefore, in need of being listed/memorized.

I argue that to understand the problematic *köpecik* and *bebecik*, we have to look at other data generally attributed to the interface of morphology and phonology. The main unit we are looking for are (weak) suppletives that can explain why the final *k* of *köpek* and *bebek* disappear when these words are suffixed. Basically, we have to explain the apparition of so-called 'stems', truncated ones (*köpe-*; *bebe-*).

If we look at Turkish phonology as a whole, there is one place where a beginning of an answer to the problem exists: again, the so-called k-Ø alternation, e.g. köpek 'dog' + -i 'acc.' → köpeği 'dog acc.' /köpei/; bebek 'baby' + -i 'acc.' → bebeği 'baby acc.' /bebei/.

The crux of the problem is again ġ. If a) we take surface forms at face value (i.e. there is no ġ; and no constraint on clusters of vowels), and b) permit paradigm interference, it becomes possible to explain how the 'truncated stems' can appear in the system. köpeġ-i /köpe-i/ and bebeġ-i /bebe-i/ permit the apparition of these 'truncated stems' (but again, only if we accept that for the native speakers, ġ is not phonemically represented and that sequences of vowels are permitted). Then, backformation/analogy can be called upon to explain the way a certain sort of paradigmatic levelling takes place. But for this, we have to accept that the full morphologically complex forms (i.e. the so-called 'outputs' of morphological processes) are somehow stored for the speaker.

9. Paradigmatic realignments/levelling

I will now discuss some cases of paradigmatic realignments/levelling that have to do with the fact that native speakers treat the so-called *ǵ* as inexistent. The result being that some morphological paradigms end up being lumped together formally, thus creating syncretism.

The traditional analysis of the written/formal form of the word *yapaca[ğ]ım* 'I will do', goes as follows:

¹² A deletion rule erasing *k* before a consonant or before *ç* would not work since *k/_C* and *k/_ç* are found everywhere in Turkish, and more importantly in the vast majority of cases involving *-çik/k_*.

- (6) yap 'to do' + aca(k) 'fut.' + m '1st pers. sg.' → yapacağım 'I will do' /yapacağım/
 (We suffix –acak to the root yap-; then suffix –m to yapacak).

Then we have two choices according to mainstream premises about Turkish phonology:

1) We say that k becomes ğ because of the buffer ı that is added between consonants – not to have *yapacakm – a form that is unpronounceable in Turkish. But this is problematic because we need to assume that there are many levels of suffixation and an ordering of these levels: yapacak-m → yapacak-ı-m → yapacağım /yapacağım/ –a) epenthesis, then b) k-∅ alternation/voicing *because* of the epenthetic vowel. But then, the constraint on the non-well-formedness of clusters of consonants, which triggered epenthesis, is now responsible for the creation of a non-well-formed cluster of vowels, by deleting k (note that k/_V is totally acceptable in Turkish).

The main problem with this analysis is still the same: we have to posit the phoneme ğ only to make our premises (i.e. invariant morphemes, no clusters of vowels) fit the data (and etymology).

2) The other option is deletion of k. This would need to be done after insertion of the epenthetic vowel – if not, -m would be inserted directly on –aca(k) when k is deleted, triggering the form *yapacam*. So epenthesis and then deletion: yapacak-m → yapacak-ı-m → yapacağım /yapacağım/ But then a paradox arises: how can a phonological rule (i.e. k deletion) create forbidden patterns (i.e. clusters of vowels)? Again, this is why ğ is needed underlyingly.

Even if these solutions were acceptable, given the current analyses of Turkish, (1-2) do not explain that what we commonly hear is *yapacağım* 'I will do'.

–m is the suffix for many morphological processes involving the first person. –m will be preceded by an epenthetic vowel if it is attached to a stem ending with a consonant, e.g. yapıyor 'he/she does' vs yapıyor-(u)m 'I do'; if the stem ends with a vowel, –m will be attached, e.g. yedi 'he/she ate' vs yedi-m 'I ate'.

The form we hear is not *yapacağım*, which could have been explained if we assume an underlying consonant (ğ) and a phonological rule of epenthesis (turning –m into –(I)m¹³). The 1st pers. suffix –m in *yapacağım* is added without epenthesis, which indicates that there is no underlying consonant that would trigger the epenthetic process.

It could also be possible to postulate a derivation adding an extra level of phonology:

- (7) yapacakm → yapacakım → yapacağım → yapacam (something we could call 'double vowel shortening' or 'vowel deletion').

The problem is that in the 2nd person singular we find *yapacan*, which contradicts this analysis: the input form is supposed to be yapacaksın → yapacan, a derivation that is unexplainable using the same type of analysis.

The only way to explain what is happening here is to look at paradigms, which can produce a –n as a 2nd person suffix:

¹³ '(I)' marks the vowel that is to be chosen in function of the phonological context (i.e. vowel harmony).

- (8) gel-di 'come + past (+Ø?) = he/she came'
 gel-di-n 'come + past + 2nd pers. = you came'

This means that the *-m* of *yapacam* and the *-n* of *yapacan* are not the result of derivations involving successive affixation followed by phonological 'arrangements'. They are the result of a merging of paradigms, probably the result of *yapacam* being treated at face value (i.e. without *ğ*), and therefore being connected to other paradigms not involving *ğ*. But for this to happen, we need a form of analogical levelling leading to syncretism, which is only possible when full forms are taken into account, may it be diachronically – through re-analysis; or synchronically.

10. Vowel Harmony

Some problems concerning Turkish Vowel Harmony (TVH) also receive a different light once we modify our statements about the Turkish vowel inventory. TVH is often blocked in the context (C)V₁CV₂, e.g. *sāt* 'watch' + *i* 'acc.' = *sāti* (**sāti*, the expected form according to the mainstream analysis of THV).

We are told it is the long vowels that block Vowel Harmony. But TVH is not always blocked by long vowels. For example, words like *sanāt* 'art' or *edebiyāt* 'litterature' are often heard as *sanāti* and *edebiyāti* in their accusative forms, thus respecting TVH, i.e. we find free variation (e.g. *sanāti* ~ *sanāti*; *edebiyāti* ~ *edebiyāti*). Another problem is that once we have said that long vowels block THV, it becomes impossible to explain e.g. *hakikat* 'truth' + *siz* '-less' = *hakikatsiz* 'without truth'; *bahs* 'topic' + 'acc.' = *bahsi*, where no long vowel is present to condition the blocking of TVH. Another case in hand is the word *yar* 'friend/lover' which used to have a long vowel, but does not nowadays. But vowel harmony is blocked as if the disappeared long vowel was still there, i.e. *yar* + *im* 'my friend/lover' (**yarim*)¹⁴.

This again points to the fact that explaining synchronic data needs to be done on its own terms. Looking at diachronic forms and rules can be helpful to understand where synchronic patterns come from. But we also have to trust the data, and not push too much analyses that involve the premise of continuity as a methodological principle if the data do not fit our predictions.

Once we accept the fact that long vowels do exist in Turkish, another picture of Vowel Harmony emerges, a picture that adds to the problems in reaching a coherent picture of this phonological rule – if we keep our premises about the relation between input and output forms intact. But no problem remains if the contemporary forms are analyzed as residues of the past, not the outputs of synchronic phonological rules. At the same time, it also becomes more difficult to posit Vowel Harmony synchronically¹⁵.

¹⁴ The examples of this type I am aware of all involve front harmony. We could eventually treat these examples as exceptions. But see Royer-Artuso ([2013]2015) for many other troubling facts about Turkish Vowel Harmony.

¹⁵ In Royer-Artuso ([2013] 2015), I show that the analysis of Turkish as having a *phonological* process of Vowel Harmony does not hold synchronically. TVH is morphological, at the least; which means that the affixes are in suppletive relations, and/or the result of analogical operations, i.e. some morphologically complex forms must be stored so that analogical processes can take place.

11. Conclusion

In this paper, I have tried to show how some of the premises we hold about Turkish impede on a realistic analysis of many related and not so related data. I have shown how these premises are interconnected, and how changing some of them have positive consequences on what we say about the phonology-morphology interface. The main conclusion is that Turkish has changed if we compare mainstream analyses to the synchronic data, that is: if we accept that the mainstream analyses are correct in their depiction of Turkish, what these analyses are analyses of is a state of the language that is now a fact of the past.

I have proposed here to take into account three of these changes, namely the analysis of modern Turkish as now 1) having long vowels; 2) having no ğ phonemically; and 3) having no constraints on clusters of vowels.

These changes might be linked to a common cause. A possible scenario could be that borrowing words with long vowels has permitted the reanalysis of some native words with ğ as including long vowels. At the same time, borrowing words with clusters of vowels has permitted to loosen Turkish's phonotactic constraints. This is at least an avenue of research worth exploring, both for the history of Turkish and in a more comparative way, i.e. looking at the way contact changes representations when words are borrowed without changing them to fit the constraints of the borrowing language.

The modifications made about Turkish phonology have consequences for the morphemic/lexical representations we adopt for this language: instead of invariant representations, (weak) suppletive ones for many exceptional patterns, but more importantly, for non-exceptional patterns as well (bringing these representations closer to the 'surface' forms).

Finally, the analyses I have presented also point to an organization of the lexicon and morphological processes that relies heavily on the paradigmatic dimension (in addition to the syntagmatic one); and point to the fact that this organization needs to be thought as based on full forms – as opposed to the mainstream analyses based on invariant morphemes + phonological rules to generate allophony and allomorphy.

References

- Anderson, S. R. 1992. *A-Morphous Morphology*. Cambridge University Press.
- Aronoff, M. 1976. *Word Formation in Generative Grammar*. Cambridge: MIT Press.
- Aronoff, M. 1994. *Morphology by Itself*. MIT Press.
- Aronoff, M. 2007. In the Beginning was the Word. *Language* 83(4): 803-830.
- Blevins, J. P. 2006. Word-based Morphology. *Journal of Linguistics* 42: 531-573.
- Blevins, J. P. 2016. *Word and Paradigm Morphology*. Oxford University Press.
- Blevins, J. 2004. *Evolutionary Phonology*. Cambridge University Press.
- Bochner, H. 1993. *Simplicity in Generative Morphology*. Berlin: Mouton de Gruyter.
- Booij, G. 2010. *Construction Morphology*. Oxford University Press.
- Bybee, J. 2001. *Phonology and Language Use*. Cambridge University Press.
- Ford, A. and R. Singh. 1991. Propédeutique morphologique. *Folia Linguistica* 25: 549-575.
- Ford, A., R. Singh and G. Martohardjono. 1997. *Pace Panini*. New York: Peter Lang.
- Göksel, A. and C. Kerslake. 2005. *Turkish: A comprehensive grammar*. Routledge.

- Inkelas, S. 2011. The Interaction Between Morphology and Phonology. In Goldsmith, Riggle and Yu (eds.) *The Handbook of Phonological Theory*. Blackwell.
- Inkelas, S. 2014. *The Interplay of Morphology and Phonology*. Oxford University Press.
- Lewis, G. 2000. *Turkish Grammar*. Oxford: Oxford University Press.
- Kabak, B. and I. Vogel. 2011. Exceptions to Vowel Harmony and Stress in Turkish: co-phonologies or prespecification? In Simon, H., Wiese, H. (Eds.) *Expecting the Unexpected: Exceptions in Grammar*, 59-94. Mouton de Gruyter: Berlin.
- Kılıç, M. A. and M. Erdem. 2008. Türkiye Türkçesindeki 'Yumuşak G' Ünsüzünün Fonetik Analizi. Proceedings of the VI Uluslararası Türk Dili Kurultayı/Ankara. 2809-2826.
- Kornfilt, J. 1997. *Turkish*. Routledge.
- Pierrehumbert, J. B. 2001. Exemplar dynamics: word frequency, lenition, and contrast. In Bybee J. and P. Hopper (Eds.) *Frequency and the Emergence of Linguistic Structure*, 137–58. Amsterdam: Benjamins.
- Pierrehumbert, J. B. 2010. *The Dynamic Lexicon*. Online version. Last consultation: 03/01/22. <http://www.phon.ox.ac.uk/jpierrehumbert/publications/DynamicLexicon.pdf>
- Pöschtrager, M. 2013. k-Ø and What Phonology Can Do. *Dilbilim Araştırmaları Dergisi* 2013/1.
- Royer-Artuso, N. 2013 [2015]. Is Turkish a Tongue-Twister? In Dziubalska-Kolaczyk, K. and J. Weckwerth (Eds.), *In memoriam of Prof. Rajendra Singh: Papers from a special session at the 44th Poznan Linguistic Meeting, September 2013*. Poznan: Adam Mickiewicz University Press.
- Royer-Artuso, N. 2016. Contact, Change and the Passage from Allomorphy to Suppletion. *Linguistic Atlantica* Vol. 35, 2.
- Royer-Artuso, N. 2017. L'Accent Lexical, l'Harmonie Vocalique et le Mot Phonologique en Turc. *Langues et Linguistique* 36.
- Stump, G. T. 2001. *Inflectional Morphology*. Cambridge University Press.
- Skousen, R. 1992. *Analogy and Structure*. Springer, Dordrecht.
- Skousen, R., Deryle L. and D. B. Parkinson (eds.) 2002. *Analogical Modeling: an exemplar-based approach to language*. John Benjamins.
- van Oostendorp, M. 2013. A Consumer Guide to Phonological Evidence. In Blaho, S., M. Krämer and B. Morén-Duolljå (eds.) *Special issue 'A Festschrift on the Occasion of X Years of CASTL Phonology and Curt Rice's Lth Birthday'*, 274-293. Nordlyd 40.1. University of Tromsø.