

Hilal Oytun Altun<http://orcid.org/0000-0003-0813-9843>

Jagiellonian University

hilal.altun@uj.edu.pl

DOI: 10.35765/pk.2022.3601.11

Modernization or Simplification: an Empirical Approach to the Turkish Language Reform

STRESZCZENIE


In this paper, I proposed an empirical approach to evaluate the accuracy of modern – or so-called purified – versions of Ottoman texts. Empirical studies on purified texts are limited in terms of quantity despite the abundance of the literature on the purification program within the Turkish language reform. I attempted to analyze a modernized text to answer the following questions; in a modernized text whether (a) semantic accuracy is preserved, or (b) the content is simplified. The method I offer here is based on a comparison of the modernized version with the authentic text from the lexical semantics point of view. My research revealed that some words with more definite meaning were replaced with more imprecise words in the modernized text. Such replacements may lead to a reduction in the content. Nevertheless, future works are needed on large scale data with the help of AI-supported tools to explore whether blurriness occurs in meaning during a modernization process.

KEYWORDS: Turkish language reform, modernization, purification, lexical semantics, simplification

STRESZCZENIE

Modernizacja czy symplifikacja: empiryczne podejście do reformy języka tureckiego

W niniejszym artykule zaprezentowano nowatorskie podejście do zagadnienia reformy języka tureckiego oparte na analizie dokładności semantycznych tekstów w języku osmańsko-tureckim i odpowiadających im tekstów w zmodernizowanym – określanym również jako rezultat procesu puryfikacji – języku tureckim. Mimo dość licznych prac naukowych poświęconych problemowi modernizacji języka tureckiego jako rezultatu przeprowadzonej w przeszłości reformy językowej liczba prac szczegółowych, których istotą byłyby badania empiryczne tekstów w języku zmodernizowanym, jest ograniczona. Analiza

Sugerowane cytowanie: Altun, H.O. (2022). Modernization or Simplification: an Empirical Approach to the Turkish Language Reform. ©  *Perspektywy Kultury*, 1(36), ss. 155–166. DOI: 10.35765/pk.2022.3601.11.

Nadesłano: 30.01.2022

Zaakceptowano: 15.02.2022

danego tekstu w języku zmodernizowanym miała na celu odpowiedź na pytanie, czy w tekście zmodernizowanym (a) została zachowana dokładność semantyczna, czy też (b) semantyka treści została uproszczona. Zaproponowana przez autorkę metoda polega na porównawczej analizie leksykalno-semantycznej tekstu w języku zmodernizowanym i jego wersji oryginalnej. W wyniku analizy okazało się, że niektóre wyrazy o specjalistycznym znaczeniu w tekście sprzed modernizacji zostały zastąpione w tekście w języku zmodernizowanym wyrazami o znaczeniu bardziej ogólnym. Zabiegi tego typu mogły w konsekwencji prowadzić do redukcji wartości semantycznej całości. Autorka podkreśla wagę badań mających na celu wykazanie, czy w procesie modernizacji języka nie doszło do redukcji lub rozmycia semantycznego, jednocześnie zauważając, że z uwagi na dużą ich skalę wykorzystanie narzędzi sztucznej inteligencji będzie nieodzowne.

SŁOWA KLUCZE: reforma języka tureckiego, modernizacja, puryfikacja, semantyka leksykalna, symplifikacja

Introduction

Turkish language reform was implemented as a part of the Westernized, modern nation-building project of the young Turkish Republic. The first step was adopting a modified Latin alphabet in 1928 which was a shock for even ideological vanguards of the nation.¹ The second and even more exhaustive step was the modernization of Turkish language. A need for reform in the written language had started to be expressed by the intellectuals from the beginning of the XIXth century (Levend, 1960, pp. 80–83). The written language of the time was criticized due to its unintelligibility caused by the complicated and long sentences stuffed with the forms copied from Arabic and Persian.

During the first decades of the republic, the language reform was initiated aiming to remove all the foreign words. Until the Third Language Congress in 1936, the plan was to purify the Turkish language via (1) reviving obsolete Turkic words, (2) collecting words from Turkish dialects, and (3) deriving new words from Turkic roots. However, this plan of relexification which was fashioned by extreme purists did not satisfy the expectations and the needs of a living language. The Congress, in 1936, where the Güneş-Dil Teorisi (Sun-Language Theory) was announced, prepared the ground in favor of a moderate approach. According to the theory, many foreign words were etymologically from Turkic, therefore it was not necessary to get rid of them (Perry, 2003, p. 248). Thus the theory functioned as

1 For the reactions to alphabet reform see: Levend, 1960, pp. 392–399.

an escape from radical purification and the commonly known loanwords in everyday language remained untouched to some degree.

The opinions on the language reform are divided over whether the Turkish language has become more modern or weaker as the consequence of the purification program, i.e. linguistic engineering harmed Turkish language or not. Advocates of both sides focus on the loanwords especially the ones borrowed from Arabic and Persian as the core element in support of their arguments. Empirical studies, on the other hand, that attempt to evaluate the effects of the reform on the language are scarce. The purpose of this article is to provide an exploratory study that seeks methods to fill this gap.

Previous works and literature

There is gigantic academic literature on the history of the reforms, debates, figures, details of the opposing and supporting ideas, etc. A quick search in Google Scholar with the “Turkish language reform” keyword in quotes gives 1300 results; adding the searching terms “purification,” “simplification,” and “debates” turns 511, 400, and 938 results respectively. To demonstrate the popularity of the topic I quote two critiques on the consequences of the so-called purification program. The first one is from a very recent newspaper article that was published in 2022:

Turkish used to be a very rich language in terms of the nuance of words, but now it has become poor, weak, stubby, and puny. Words that existed until 40–50 years ago, each of which has a different nuance, are now usually confined to a single newly invented word. Turkish is advancing at full speed to become a language of 150–200 words, spoken by African natives. It even already happened (Bardakçı, 2022, <https://www.haberturk.com/yazarlar/murat-bardakci/3320559-davutoglu-kadm-dil-diyor-ama-kurce-kadm-yani-olu-degil-yasayan-bir-dildir>).

The second quote is from the foreword of one of the modern standard Turkish dictionaries:

What drives us to this path [to write a dictionary – H.O.A.] is the fact that our language is getting poorer and losing its ability of articulation as a result of deliberate interventions to the Turkish language. It is worth considering how the 12 words that were used before, such as ‘aşikâr’ (overt, apparent), ‘bedîhî’ (self-evident), ‘dekolte’ (décolleté), ‘münhal’ (vacant), and ‘müstehcen’ (obscene), are met with a single word: ‘açık’ (open),

confuses the language, how the nuances are lost and how this intervention impoverishes the language! (Ayverdi, 2010, p. VI).

There is an ambiguity regarding the terminology in Turkish academic literature. The term ‘simplification’, by definition, is employed in the context of language learning. In some other fields of study though, it often appears with the meanings of ‘nativization/purification/Turkification’ in the sense of ‘modernization’ (Bulut, 2014). Needless to say, the term ‘simplification’ itself implies a reduction in content. The simplification process aims to make a text accessible for the people with specific needs; e.g. for people with aphasia or dyslexia, lay readers of technical texts and low literacy readers, etc. However, the target group of the modernized texts is much more different than that. Whereas simplification is the task of diminishing the complexity of a text, the purpose of modernizing a text is to increase the understandability of an old text by contemporary readers. In this study I preferred the term ‘modernization’ since the language program implemented on Turkish neither a ‘simplification’ nor a real ‘purification’.

A language reform is a process that is inescapably contaminated by a combination of extra-linguistic factors. More specifically: “linguistic engineering is a perilous branch of socio-political experimentation ... [and – H.O.A.] has been practiced mainly not by linguists but generals, politicians, social ideologues and other amateurs” (Perry, 2003, p. 238). This particularity of the topic makes it attractive for the diverse fields of study.² The debate on the language reform, publications, and the ideological ground of the conflicting ideas as well as the implementation process of the reforms are among the popular topics of history studies. Linguists, on the other hand, focus on the lexical changes, especially on the suggested Turkish replacements for loanwords (Bayar, 2003; Timurtaş, 1979).

Modernization of Ottoman texts recently emerged as a topic in machine translation studies as well (Kurt & Bilgin, 2012). The developments in optical character recognition (OCR) field also stimulated computer engineers’ attention to the topic. Converting scanned Ottoman material to machine-encoded texts is a relatively new research field. The ideal application would take scanned images as input and give a modern Turkish version of the text as output to fulfill the needs of a random reader who isn’t familiar with Ottoman. Three consequent steps are required for a successful outcome: (1) OCR on scanned image, (2) transcription, and (3) translation (Dölek & Kurt, 2021). Although a couple of OCR tools with high accuracy levels are available, automatic Latinization of Arabic scripted Ottoman texts is currently far from meeting expectations. Automatically

2 For a comprehensive bibliography see: Demircan & Erözden, 1992.

transcribed Ottoman texts can be read on the digital collections of IRCICA (<https://library.ircica.org/Pages/Collections>), but the accuracy of the transcription is fairly poor and similar accuracy rates are also valid for ottoman.com, which is a relatively successful application in terms of OCR.

There are also found a couple of machine translation researches from Ottoman to modern Turkish (Bakırcı, 2019) and they point to the need for larger datasets (Özkan, 2018, p. 62). Building and annotating an electronic parallel corpus of modernized and authentic texts appears to be the primary need. Likewise, to analyze repetitive patterns automatically is dependent on computational research on large scale data.³

To conclude this section, the following opinion is offered as a core concept for future research:

Research in LPP [language policy and planning] must be understood as both a multidisciplinary and an interdisciplinary activity, in that conceptual and methodological tools borrowed from various disciplines need to be appropriately integrated and applied to real-world problems and challenges involving language, which, by definition, are embedded in all aspects of society and social life (Ricento, 2006, p. 9).

Modernization process of a text

The texts I chose for my research are the second Constitution which was accepted in 1924 and its modernized 1945 version.⁴ What makes this text unique is that the modernized version has official acknowledgment for the equity of its content to the authentic text. The Grand National Assembly, accepted the 1945 Constitution as the exact equivalent of the 1924 Constitution (Karlıklı, 1999, p. 57). There are numerous modernized texts of the then literature and historical works, but none of them has such approval. For example, there are a dozen of modernized, simplified, shortened and purified versions of *Nutuk* (The Speech) by Atatürk,⁵ but none of them was evaluated in terms of authenticity (Korkmaz, 2004, pp. IX–XII).

The 1924 Constitution text – and its 1945 version – is a brief and concise text that is convenient for manual comparative lexical analysis. There

3 Studies on automatic analysis of simplified texts (Crossley et al., 2007) and automatic simplification systems (Shardlow, 2014) have been conducted for more than a decade. When it comes to Turkish, as I saw from the literature, similar studies on modernized Turkish texts haven't started yet.

4 Full texts of the 1924 and 1945 Constitutions are available at: <https://www.anayasa.gov.tr/tr/mevzuat/onceki-anayasalar/1924-anayasasi/>

5 For the debate on the purification/modernization of the *Nutuk*, see: Uzun, 2005, pp. 116–123.

are found 2624 words in total, 797 unique words in 1924 text, and the numbers decrease to 2407 and 651 respectively in its modernized 1945 version. To be able to list all the instances that differ in the two texts I used a text comparison tool available as a web application (https://neil.fraser.name/software/diff_match_patch/demos/diff.html). Then I listed the unique words and grouped them according to their origins. The distribution of the words according to origin is as follows: The 1924 text contains 670 borrowings, 112 words from the Turkic origin, 9 compound words whose components are from diverse origins, and lastly 5 proper names. While the number of loanwords in the 1945 modernized version is 166, the number of Turkish originated words increases to 472.⁶

Modernization of the text is realized through (1) grammatical, and (2) lexical replacements. Grammatical replacements occur in two distinct levels, (a) morphological level, and (b) word-order, i.e. syntactic level. The purpose of the grammatical changes is to replace the difficult forms with more familiar substitutions. Thus the publicly comprehensible forms are preferred in the modernized text and the origin of the word stem is ignored in most of the cases.

In group (1a) that includes morphological changes, grammatical forms, borrowed from Arabic and Persian are substituted with Turkish suffixes.

memuriyet > memurluk (service, official duty) (< Ar. ma‘mūr)
sahtekârlık > sahtecilik (forgery) (< Pers. sāhta)
mahakim > mahkemeler (courts) (< Ar. maḥkamat)
mezun > izinli (licensed/approved) (< Ar. izn)

Substituting prepositions and grammatical words with Turkish suffixes causes change in word order:

bilâ kaydü şart (without reserve and condition) > kayıtsız şartsız
(unconditionally)
adem-i devam (absence of attendance) > devamsızlık (nonattendance)

Group (1b) includes the phrases whose syntactic structures borrowed from Persian or Arabic. In some instances, the structure of the phrase is changed, and the borrowings in it are left untouched.

ahkâm-ı esasiye > esas hükümler (fundamental provisions)
kuva-yı harbiye > harb kuvvetleri (armed forces)
cezaî bir hüküm > bir ceza hükmü (a penal sentence)
şekl-i devlet > devlet şekli (form of state)

6 For the lexical statistics of the Constitutions of Turkey in terms of origin throughout the time, see: Özdemir, 2003, pp. 149–155.

Some changes occur due to the differences between verb structures of Arabic and Turkish:

intihap edilmeleri caizdir [election do/make-PAS-PL-PSS permissible]
(it is permissible for them to be elected) > seçilebilirler [elect-PAS-MOD-
-PRS-PL] (they can be elected)

Like the example above, rewriting can cause lexical changes:

her türlü müdahalâttan âzade ol- (to be free from any kind of interven-
tion) > hiçbir türlü karışılama- (cannot be interfered with any means)

In some instances where the content words are deleted in the purified version, and their functions are substituted by syntactic arrangements:

bir Türk babanın sulbünden doğan (the one who was born from a Turkish
father's progeny) > bir Türk babadan gelen (the one who is coming from
a Turkish father)
bir zat uhdesinde (in or under a person's charge/responsibility) > bir
kişide (upon a person)

In lexical replacements, which I put in group (2), the origins of the words are the first matter of importance. This kind of replacements serves to Turkification of the text. Turkification of the terminology is also one of the most argued themes related to language reform (Safa, 1970, pp. 25–26, 52, 64–72). Since the terms are the words that are naturally bound with a specified and predefined semantic domain, they usually appear as loan translations. A couple of examples from the terminology are:

hükmî şahsiyet > tüzel kişilik (legal entity)
kuvve kazaiye > yargı erki (judiciary)
ekseriyeti mutlaka > saltçokluk (absolute majority)
akit > bağit (contract)

There are other calques found in the text other than the terminology:

fevkalâde > olağanüstü (extraordinary)

Even though lexical replacements are Turkifications most of the time, there are some borrowings that are substituted with other borrowings:

evrak (< Ar. avrāk) > kâğıtlar (< Pers. kāğid) (documents, papers)
derhal (< Pers. darhāl) > hemen (< Pers. hamān) (immediately)

It is difficult to argue that there is a semantic change, i.e. content reduction, in the forms I explained above. To be able to find out whether a reduction occurred in content it is needed to examine the substitutions according to their semantic domains; for example, replacing the words with more generic ones may cause blurriness in meaning.

Restriction of usage is an indicator for the content bearing potential of a token, as well. According to the theory, for a linguistic unit to transmit information, the probability of being in a certain place in the discourse should be less than 1 and greater than Null. As the probability goes near to 1/1 the content of the unit decreases (Gemalmaz, 1992, p. 170). In other words, more restricted linguistic units carry more information, e.g. specific words are more restricted than generic words. I have examined the purified text from this restriction theory point of view. I have checked the frequency of the given words in Turkish corpus (Sezer & Sezer, 2013). As the polysemy of a word is another indicator for frequency, I have referred to the modern standard Turkish dictionaries (sozluk.gov.tr, lugatim.com) as well as *Redhouse's Turkish and English Lexicon* (1890).

In the modernized text, a couple of words are employed more frequent as a substitution. For example, the frequency of 'yap-' (make/do) is 23 instances in 2407-word text. The verb 'yap-' is, as might be expected, one of the most generic verbs in Turkish. The words that were substituted with 'yap-' in the authentic text are as follows:

akd (constitute and open -a meeting), icra (put into execution), ifa (fulfill, perform), tanzim (organize, put in order)

Another frequent word 'aykırı' (contradictory) substitutes three distinct words:

mugayir (opposed, contrary, adverse to), muhalif (opposing, opposed, contrary to), münafi (that excludes, incompatible, irreconcilable)

The word 'gerek' (necessary) is also frequently employed one, in adjective 'gerekli' or verb 'gerek-' forms, and its counterparts are 'icap', 'lüzum', 'iktiza' (requisite, necessary, requirement, necessity). Those three words usually appear in similar contexts and don't express very distinct meanings, therefore substituting those three words with 'gerek' can be seen as a style issue rather than a semantic one.

There are other instances where one Turkish word was employed to substitute multiple loanwords. Some examples are:

toplan- (collect, gather, etc.) < temerküz (concentration), içtima (gathering, meeting)

konut (housing) < mesken (a dwelling, a home), ikametgâh (place of residence)
al- (take) < cibayet (tax, tribute), ittihaz (taking to oneself), istimlak (expropriation)
yetki (authority) < mezuniyet (authority), salahiyet (authority)
kaldır- (remove) < ıskat (dismissing from office), fesih ve ilga (annuling and abolition)

Some borrowings in authentic text are replaced despite the fact that they are publicly known and used widely. The substitutions are either more generic words or newly invented ones. The frequency numbers are taken from tscorpus.com.

cereyan et- (take place, happen, occur) (frequency: 4.26 instances per million words) > yürü- (to go forward) (frequency: 237.19 instances per million words)
idare et- (to manage) (frequency: 27.52 instances per million words) > yürüt- (to carry out) (298.26 instances per million words)
davet (to invite) (frequency: 112.00 instances per million words) > çağır- (to call) (frequency: 81.75 instances per million words)
ödev (duty) (frequency: 2.28 instances per million words) < mükellefiyet (liability) (frequency: 0.30 instances per million words), mecburiyet (obligation) (frequency: 0.81 instances per million words)

As it can be seen from the numbers, ‘yürü-’ and ‘yürüt-’ are much more frequent than the words they substitute. This is an unmistakable indicator for a generalized semantic domain.

Another replacement type is that two words with close meanings are substituted with one and usually more generic word in modernized version:

hakiki ve yegane (the true and the sole) > ancak (the only)
tesbit ve tayin (fixing, demonstrating and appointing) > çiz- (draw)
fesih ve ilga (annuling and abolition) > kaldır- (remove)

In order to avoid a vagueness that can be caused by the generic content of the substitutions auxiliary elements are added:

hasrı nefis etmekten ayrılma-... (not give up dedicating oneself) > olanca varlığımla çalışmaktan asla ayrılma-... (never give up working with my whole existence)
hiçbir fedakârlık yapmağa zorlanama-... (cannot be force to do any sacrifice) > hiçbir şey yapmaya ve vermeye zorlana-... (cannot be forced to do and give anything)

There can be observed instability in some cases where a certain word was replaced with various substitutions in modernized text. For example, 'istimlak' (expropriation) is substituted with the newly invented 'kamulaştır-' (expropriate) and 'al-' (take) is employed in modern version. 'al-', like the 'yap-' (make/do), 'yürü' (go forward) above, is a one of the most frequent generic verbs in Turkish. Other examples of this type from the text are:

inhilâl (vacant) > açık (open), boş (empty)
riayet (respect) > say- (respect), uy- (compliance)

Conclusion

In the present study, I investigated one of the consequences of the process of modernization, i.e. reduction in content. I compared the modernized version and authentic text of 1924 Constitution of Turkey from the lexical semantics perspective. Findings show that the modernization process resulted in a reduction in the content to some extent especially due to the overuse of generic verbs. Nevertheless, deciding whether the findings are valid for modernized texts in general, requires further research on more representative data.

ABBREVIATIONS

Ar.	Arabic
MOD	Modal
PAS	Passive
Pers.	Persian
PL	Plural
POS	Possesive
PRS	Present tense

BIBLIOGRAPHY

- Ayverdi, İ. (2010). *Kubbealtı Lugatı, Misalli Büyük Türkçe Sözlük*. İstanbul: Kubbealtı.
- Bakırcı, A. (2019). *A Deep Learning Based Translation System from Ottoman Turkish to Modern Turkish*. Unpublished master thesis. İstanbul: Gebze Technical University.

- Bardakçı, M. (2022). *Davutoğlu “kadim dil” diyor ama Kürtçe kadim, yani “ölü” değil; yaşayan bir dildir!* Retrieved from: <https://www.haberturk.com/yazarlar/murat-bardakci/3320559-davutoglu-kadm-dil-diyor-ama-kurtce-kadm-yani-olu-degil-yasayan-bir-dildir> (access: 23.01.2022).
- Bayar, N. (2003). *1930’lu yıllardan günümüze kadar Türkçe’de yeni türetilen kelimelerin yapımı ve dil bilgisi bakımından değerlendirilmesi*. Unpublished PhD thesis. İstanbul: Marmara University.
- Bulut, M. (2014). Atatürk’ün Türkçeye Yönelik Özleştirme/Sadeleştirme Çalışmaları ve Bu Bağlamda Yaşanan Dil Tartışmaları Üzerine Bir Değerlendirme. *Electronic Turkish Studies*, 9 (11), 131-147.
- Crossley S.A., Louwse M.M., McCarthy P.M., & McNamara D.S. (2007). A Linguistic Analysis of Simplified and Authentic Texts. *The Modern Language Journal*, 91 (1), 15–30. DOI: 10.1111/j.1540-4781.2007.00507.x. Retrieved from: https://www.researchgate.net/publication/227718830_A_Linguistic_Analysis_of_Simplified_and_Authentic_Texts (access: 13.01.2022).
- Demircan, Ö. & Erözden, A. (1992). Yazı Devrimi Kaynakçası. *Dilbilim Araştırmaları Dergisi*, 3, 117–134.
- Dölek, İ. & Kurt, A. (2021). *Ottoman OCR: Printed Naskh Font*. International Conference on INnovations in Intelligent SysTems and Applications (INISTA). DOI: 10.1109/INISTA52262.2021.9548616.
- Gemalmaz, E. (1992). Türkçenin Morfo-sentaktik Yapısının Fonolojisine Etkileri. *Türk Dili Araştırmaları Yıllığı – Belleten*, 40, 169–173.
- Karlıklı, Y. (1999). Türk Anayasaları ve Dil. *İ.Ü. Siyasal Bilgiler Fakültesi Dergisi*, 20, 56–60.
- Korkmaz, Z. (2004). *Kemal Atatürk Nutuk 1919–1927*. Ankara: Atatürk Kültür ve Tarih Yüksek Kurumu.
- Kurt, A. & Bilgin, E.F. (2012). The Outline of an Ottoman-to-Turkish Automatic Machine Transliteration System. *First Workshop on Language Resources and Technologies for Turkic Languages*, 45–50, İstanbul, Turkey, 10–15 October.
- Levend, A.S. (1960). *Türk Dilinde Gelişme ve Sadeleşme Evreleri*. Ankara: Türk Dil Kurumu.
- Özdemir, H. (2003). *Türkçenin Sadeleşme Tarihinde Anayasaların Dili*. Unpublished master thesis. Ankara: Ankara University.
- Özkan, E. (2018). *Eski Türkçe Metinlerin Günümüz Türkçesine Sadeleştirilmesi*. Unpublished master thesis. Ankara: Hacettepe University.
- Perry, J. (1985). Language Reform in Turkey and Iran. In: T. Atabaki & E.J. Zürcher (eds.), *Men of order: Authoritarian Modernization in Turkey and Iran*. New York: I.B. Tauris, 238–259.
- Redhouse, J.W. (1890). *A Turkish and English Lexicon*. Constantinople: A.H. Boyajian.

- Ricento, T. (2006). Theoretical Perspectives in Language Policy: An Overview. In: T. Ricento (ed.), *An introduction to language policy: theory and method*. Blackwell Publishing, 3–9.
- Safa, P. (1970). *Osmanlıca, Türkçe, Uydurma*. İstanbul: Ötüken.
- Sezer, B. & Sezer, T. (2013). *TS Corpus: Herkes İçin Türkçe Derlem*. Proceedings 27th National Linguistics Conference. May, 3–4 Mayıs 2013. Antalya, Kemer: Hacettepe University, English Linguistics Department, 217–225.
- Shardlow, M. (2014). Out in the Open: Finding and Categorising Errors in the Lexical Simplification Pipeline. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'12)*. Reykjavik, Iceland, May. European Language Resources Association (ELRA), 58–70. Retrieved from: http://www.lrec-conf.org/proceedings/lrec2014/pdf/479_Paper.pdf (access: 22.01.2022).
- Timurtaş, F.K. (1979). *Uydurma Olan ve Olmayan Yeni Kelimeler Sözlüğü*. İstanbul: Umur Kitapçılık.
- Uzun, H. (2005). *Atatürk'ün Nutuk'unun İçerik Analizi*. Unpublished PhD thesis. Ankara: Hacettepe University.

Online sources

- 1924 Anayasası <https://www.anayasa.gov.tr/tr/mevzuat/onceki-anayasalar/1924-anayasasi/>
- Diff, Match and Patch https://neil.fraser.name/software/diff_match_patch/demos/diff.html
- IRCICA Farabi Digital Library <https://library.ircica.org/Pages/Collections>
- Kubbealtı Lugatı <http://lugatim.com/>
- Osmanlıca OCR <https://www.osmanlica.com/>
- TSCorpus <http://cqpweb.tscorpus.com/cqpweb/>
- Türk Dil Kurumu Sözlükleri <https://sozluk.gov.tr/>

Hilal Oytun Altun – Ph.D. in Turkic philology, Marmara University (İstanbul), 2009. She works primarily on linguistic features of Ottoman, contemporary Turkish, and Kazakh languages with a particular focus on inter-lingual relations. She is a member of Turkology department at the Institute of Oriental Studies at the Jagiellonian University. She is the author of the book: *Tercüme-i Tevârih-i Mülûk. Nevâyî'nin Fars Hükümdarları Tarihi* [Translation of History of Dynasties. Nevâyî's History of Persian Dynasties] (2020) which was awarded with the “Nagroda Rektora indywidualna III stopnia za osiągnięcia naukowe” by Jagiellonian University in 2021.