

A Comparative Linguistic Analysis of Ottoman and Modern Turkish through UD Treebanks

Enes Yilandilođlu

University of Helsinki
Metsätalo, Unioninkatu 40, 00170 Helsinki, Finland
enes.yilandiloglu@helsinki.fi

Abstract

While the linguistic shifts between Ottoman and modern Turkish are well-documented qualitatively, quantitative analyses remain scarce. This study addresses this by conducting a comparative computational analysis using two Universal Dependencies treebanks: OTA-DUDU for Ottoman Turkish and TR-BOUN for modern Turkish. By employing descriptive statistics and a log-likelihood ratio test, we demonstrate the change and quantify the magnitude of diachronic variation. The analysis yields three primary statistical findings. First, our data reveals a 77% compliance rate with labial vowel harmony for suffixes, while this value is 98% in modern Turkish. This discrepancy can be explained by the presence of rounding in Ottoman Turkish, which disappears in modern Turkish. On the other hand, the compliance rate of palatal vowel harmony is quite high for both languages, 96% for Ottoman Turkish and 99% for modern Turkish. Second, some suffixes, such as the converb $-(y)lp^1$ and the dative infinitive $-mAyA$, changed by reducing their allomorphs in modern Turkish. Third, we demonstrate that Arabic and Persian pluralization rules, which constituted 28% of plural nouns in Ottoman Turkish, lost their pluralizing function in modern Turkish, although the words remain with singular meaning.

Keywords: Universal Dependencies, Ottoman Turkish, modern Turkish, linguistic variation

1. Introduction

The transition from Ottoman Turkish to modern Turkish involved not only an alphabet change but also significant lexical, phonological, and morphological shifts. Although such changes were documented in previous qualitative studies, the use of quantitative methods remained limited. This study differentiates itself by utilizing annotated Universal Dependencies (UD) treebanks, OTA-DUDU (Yilandilođlu and Siewert, 2025) for Ottoman Turkish and TR-BOUN (Marşan et al., 2022) for modern Turkish, to conduct a comparative analysis. It demonstrates that such changes can be traced through treebanks to follow how Turkish evolved throughout time. Thus, it analyzes historical linguistic changes in terms of vowel harmony in suffixes attached to nouns, the change in certain suffixes, and pluralization rules. First, Ottoman Turkish exhibited a tendency toward rounding in vowels (labialization). In contrast, modern Turkish enforces stricter labial vowel harmony for grammaticality, transforming Ottoman forms like *insānuñ* (human-2SG.POSS) into *insanın* (human-2SG.POSS). Despite these shifts, certain exceptions persist in modern Turkish due to loanwords like *bahsi* (topic-ACC) and *halim* (condition-1SG.POSS). Second, the changes in the suffixes, $-(y)lp$ and $-mAyA$ were further investigated, revealing the obsolescence of certain allomorphs. Third, it shows how pluralization takes place in Ottoman and modern Turkish.

2. Background

Turkish is a language in the branch of the Altaic language family (Lewis, 1977, v). Ottoman Turkish refers to the historical version of Turkish (see Demir and Yılmaz, 2002 for a discussion on naming Ottoman Turkish and Kerlake, 2015, 180-1 for problems of terminology).

The vowel harmony in modern Turkish is a phonological process to determine the vowels in a word except the first vowel (Göksel and Kerlake, 2005, 21). Vowel harmony in both Ottoman Turkish and modern Turkish has two subsystems: palatal harmony and labial harmony. Palatal harmony (2-way alternation) is related to the selection of low vowels in suffixes: back vowels (a, ı, o, u) trigger $-a$, while front vowels (e, i, ö, ü) require $-e$. Labial harmony (4-way alternation) is for the selection of high vowels in suffixes based on both backness and rounding: back unrounded vowels (a, ı) trigger $-ı$, front unrounded (e, i) trigger $-i$, back rounded (o, u) trigger $-u$, and front rounded (ö, ü) trigger $-ü$ (see Clements and Sezer 1982; Lewis, 1977, 19-22; Özçelik 2024, 128-157 for detailed explanation of vowel harmony in Turkish). Yet the main difference between Ottoman Turkish and modern Turkish lies in the labial harmony. Especially during the early stages, Ottoman Turkish allowed the forms like *gelüp* (come-CVB) or *vilāyetinüñ* (province-3SG.POSS-n-GEN) in its suffixes where the suffix can become rounded, even though the palatal harmony is maintained. Modern Turkish, conversely, is stricter in such cases and requires labial harmony for

¹ In this paper, A refers to low/unrounded vowels for palatal harmony (a and e), I refers to high vowels for labial harmony (ı, i, u, and ü), B refers to b and p, and G refers to g and ğ.

grammaticality, changing *gelüp* (come-CVB) to *gelip* (come-CVB) and *vilâyetinüñ* (province-3SG.POSS-n-GEN) to *vilâyetinin* (province-3SG.POSS-n-GEN). However, a few exceptions remain in modern Turkish due to loanwords, such as *bahsi* (topic-ACC) and *halim* (condition-1SG.POSS) (Göksel and Kerslake, 2005, 24).

Studies (Demir and Yılmaz, 2002, 853; Timurtaş, 1976, 332, 334) argue in Ottoman Turkish and Old Anatolian Turkish, a period encompassing Ottoman Turkish between the 14th and the 15th centuries (Demir and Yılmaz, 2002, 853), vowel harmony had a tendency towards rounding in vowels (labialization), and labialization is observed in the vowels of many suffixes. Similarly, Stachowski (2013, 551) analyses the labiality of *-lık* suffix, and points out the presence of “dislabial” vowels in Ottoman Turkish. It was also stated that labial vowel harmony became prevalent in Ottoman Turkish before modern Turkish after the 17th century (Develi, 1995; Gümüşkılıç, 1997; Kerslake, 2015, 185). On the other hand, palatal harmony was well-established in Ottoman Turkish (Demir and Yılmaz, 2002, 861). Additionally, Timurtaş (1976, 341) considers *b/p* change as another difference between Old Anatolian Turkish and modern Turkish while Kerslake (2015, 185-6) argues in modern Turkish, stops in word-final position have become unvoiced (devoiced). The converbial suffix *-(y)lp* was also realized as *-(y)lBAn* which was not encountered in Turkish grammar (Timurtaş, 1976, 336). The converb *-(y)lp* was also used to link clauses in Ottoman Turkish sentences (Kerslake, 2015, 198).

Another suffix *-mAK* followed by the dative suffix is realized as *-maya* and *-meye* in modern Turkish, instead of *-mağa* and *-meğe* which are the older forms (Göksel and Kerslake, 2005, 87). Alpay (2015, 12) also emphasizes the use of soft *g* (*ğ*) in such circumstances as an old variation. However, the old variation was still documented in Turkish (Lewis, 1977, 84).

Lastly, pluralization in Ottoman Turkish consists of Turkish, Arabic, and Persian grammatical rules. While modern Turkish pluralization relies on the agglutinative suffix *-lAr* (*-lar* or *-ler* depending on vowel harmony), Ottoman Turkish utilized pluralization rules from Arabic and Persian. From Arabic, words were pluralized by using suffixes, such as *-în*, *-ûn*, and *-ât*, and by changing the internal structure of a singular word, known as broken plurals (*cem-i mükesser*) for instance, from *fikr* (idea) to *efkâr* (idea.PL). From Persian, on the other hand, the suffixes *-ân* and *-hâ* were also used. It should be noted that both Arabic and Persian words can be pluralized through the Turkish suffix, *-lAr*. In contrast, modern Turkish simplified this complex system by standardizing the use of *-lAr* suffix for all nouns.

Various studies in the domain of historical linguistics analyzed language variation across time with the support of computational methods (e.g., Nevalainen et al., 2011 for English and Montes et al., 2024 for Spanish). Such studies also utilized parsed treebanks for the data such as Eckhoff et al. (2020). When it comes to the change from Ottoman Turkish to modern Turkish, studies such as Cristaldi (2025) and Petrova and Kuleshova (2020) used dictionaries as the main data source. Other studies leveraged primary sources for the comparison (e.g., Yağmur, 2025). This study differentiates itself from others in the Turkish domain by utilizing annotated treebanks as the data and leveraging computational methods to analyze the change.

3. Data and Methods

This study utilized two treebanks in Universal Dependencies, namely OTA-DUDU for Ottoman Turkish, the largest treebank for the language and TR-BOUN for modern Turkish. OTA-DUDU was chosen as it covers various genres from the 14th to the 20th centuries, though the data is not equally distributed across the centuries, while TR-BOUN was selected since its annotation guidelines align with OTA-DUDU the most among ten Turkish treebanks. The statistics of the OTA-DUDU and TR-BOUN can be seen in Table 1.

	Total tokens	Total words	Total sentences
OTA-DUDU	16,728	17,125	1,782
TR-BOUN	121,835	125,212	9,761

Table 1: Overall statistics of the treebanks

As shown in Table 1, TR-BOUN is considerably larger than OTA-DUDU. We did not downsample TR-BOUN to match the Ottoman corpus size because the G^2 log-likelihood test handles unequal corpus sizes. Absolute frequency comparisons, however, are not normalized and should be interpreted with this imbalance in mind.

Subsequently, we normalized each token by removing diacritics for specific characters (e.g., $\tilde{n} \rightarrow n$, $\dot{g} \rightarrow g$). This step was essential as the target treebanks utilize different transliteration schemes, namely Ottoman Turkish transliteration chart (Birnbäum, 1967) and modern Turkish alphabet (Lewis, 1977, 11). This study analyzes three phenomena: suffix variation, vowel harmony compliance, and pluralization patterns. The variations can be seen in Table 2.

Feature	Old variation	Incoming variation
Converb	<i>-(y)lB(An)</i>	<i>-(y)lp</i>

Dative infinitive	-mAGA	-mAyA
Vowel harmony-palatal during suffixation	Strict 2-way harmony (Front/Back)	Strict 2-way harmony (Front/Back)
Vowel harmony-labial during suffixation	Partial Inconsistent (Generalized Rounding)	Standardized / Strict (4-way harmony)
Pluralization	Turkish/Arabic/ Persian pluralization rules	Only Turkish plural rule

Table 2: Variations

The method consists of statistical analysis of the treebanks in a comparative way. In addition to applying the techniques in descriptive statistics such as absolute frequency, this study also leverages log-likelihood ratio test (G^2), a method proposed by Dunning (1993) and previously used in computational and historical linguistics (Rayson et al. 2004 and Lijffijt et al. 2012, also see Lijffijt, 2013 for potential issues in this method). The method is resilient to data imbalance. Log-likelihood ratio test was calculated via the formula below.

$$G^2 = 2 \sum O_i \ln \left(\frac{O_i}{E_i} \right)$$

In this formula, O_i denotes the observed frequency of a feature in corpus i (OTA-DUDU or TR-BOUN), and E_i refers to the expected frequency under the null hypothesis of no difference between corpora. The summation is performed over all four cells of the contingency table (feature present/absent \times Ottoman/modern Turkish).

The analysis was conducted on token counts normalized via log-likelihood, which is robust to large corpus size asymmetries. This approach assesses not only the statistical significance but also the magnitude and direction of change. By doing so, it aims to present what changes and remains in the languages in a quantifiable way.

4. Results

This section presents the results of this study in three subsections: suffix variation analysis, vowel harmony compliance, and pluralization patterns.

4.1 Suffix Variation

To analyze the data, we then explored linguistic variation by two suffixes. These are (i) verb forms ending in $-(y)lB(An)$ in Ottoman Turkish vs. $-(y)lp$ in modern Turkish and (ii) verbal nouns ending in

$-mAGA$ ($-ma\acute{g}a/-me\acute{g}e$) versus the common version in modern Turkish $-mAyA$ ($-maya/meye$), known as dative infinitive. The first suffix derives adverbs from verbs, the second derives nouns from verbs. They were detected using a rule-based approach: UPOS is VERB, and the suffixes end with either $-(y)lB(An)$ or $-mAGA/-mAyA$. Here the aim was not to extract pairs but to acquire all instances in each treebank.

Table 3 provides the frequency distribution for each variation across treebanks.

	Ottoman Turkish	Turkish
VERB-mAGA	28	0
VERB-mAyA	16	648
VERB-(y)lBAn	0	0
VERB-(y)lb	7	0
VERB-(y)lp	453	558

Table 3: Frequency table for linguistic variations

Table 3 demonstrates the obsolescence of $-mAGA$ and $-(y)lb$ in modern Turkish while they were employed in Ottoman Turkish. It also shows $(y)lBAn$ was completely absent in both treebanks.

4.2 Vowel Harmony Compliance

We analyzed the shift through vowel harmony compliance on two different subsets. The first subset consists of all nouns with suffixes in both treebanks, yielding 3,456 Ottoman Turkish tokens and 21,500 Turkish tokens. This subset provides the overall result; however, it has words that do not appear in the other treebank which results in an imbalance in the lexical coverage. The second subset includes identical noun pairs. The subset was prepared by using a rule-based string-matching approach where the normalized form starts with the normalized lemma. With this approach, we aimed to reduce the impact of lexical variation. We identified 226 shared lemmas across both corpora, which gives 217 identical lemma-feature combinations. This resulted in 429 Ottoman tokens and 1,866 Turkish tokens. The pairs were then examined to quantify historical shift in terms of vowel harmony. Hereafter, the first subset will be referred to as *all nouns* and the second one will be called *matched pairs*.

By utilizing these two subsets, tokens were selected based on whether their suffix contains the relevant vowel type for each harmony. For palatal harmony, only tokens whose suffix contains a low vowel (a or e) were analyzed, as these are the vowels that alternate under palatal harmony. For labial harmony, only tokens whose suffix contains a high vowel (i , $ı$, u , or $ü$) were analyzed, as these are the vowels that alternate under labial harmony. Tokens whose suffix lacks

the relevant vowel type are ignored for each analysis, as they provide no evidence for or against the harmony rule. This filtering is based on a rule-based approach to detect the noun suffixes first, excluding the characters except the lemma and identifying the words that are eligible for each vowel harmony analysis.

Table 4 presents the rates for vowel harmony compliance. It shows that palatal vowel harmony compliance is high in both treebanks, while labial harmony shows a prominent difference between Ottoman Turkish (77.06%) and modern Turkish (98.42%) when all nouns are considered. A similar view can be observed while only identical pairs are analyzed. While palatal vowel harmony is 100%, labial vowel harmony in Ottoman Turkish is lower than labial vowel harmony in modern Turkish. An example for the change in labial harmony is *öfiine* (front-3SG.POSS-DAT) in Ottoman Turkish and *önüne* (front-3SG.POSS-DAT) in modern Turkish.

Type of Vowel Harmony	Ottoman Turkish – all nouns (%)	Modern Turkish – all nouns (%)	Ottoman Turkish – matched pairs (%)	Modern Turkish – matched pairs (%)
Palatal	96.04 (n=1,941)	98.75 (n=14,141)	100 (n=313)	100 (1,460)
Labial	77.06 (n=2,568)	98.42 (n=13,959)	81 (n=321)	99.85 (1,317)

Table 4: Vowel harmony compliance rates

Additionally, we calculated the use of variations via log-likelihood ratio, which can be seen in Table 5.

Feature	G ²	p-value
Dative Infinitive (-mAGA vs. mAyA)	154	<0.001
Converb (- (y)IB(An) vs. - (y)Ip)	12	<0.001
Palatal Vowel Harmony – all nouns	61	<0.001
Labial Vowel Harmony – all nouns	1,431	<0.001

Table 5: Log-likelihood ratios

Table 5 presents statistically significant results for all linguistic variations. Firstly, it demonstrates a statistically significant change in terms of the dative infinitive, -mAyA, in Turkish. Secondly, it provides a statistically significant shift to the converbial suffix -(y)IB ending with *p*. Lastly, it shows statistically significant result for palatal and labial vowel harmony. Moreover, the G² values indicate the magnitude of difference between the two corpora: higher values represent greater divergence. Notably, labial vowel harmony shows the highest G² value (1,431), indicating that this feature exhibits the most substantial difference between Ottoman and Modern Turkish among all variables. However, palatal vowel harmony shows a considerably lower G² value (61), suggesting less divergence compared to labial vowel harmony. Among suffix variations, the dative infinitive (-mAğA/-mAyA) shows higher divergence (G²=154) than the converb suffix (G²=12).

Further analysis reveals that while vowel harmony, in general, became more restricted, some irregularities persist in modern Turkish, such as *halinde* (state-3SG.POSS-n-LOC) and *saatiyle* (time-3SG.POSS-y-INS) which prevents the vowel harmony compliance rate from reaching 100%.

4.3 Pluralization Patterns

Lastly, we categorized plural nouns based on how they are pluralized: the standard Turkish suffix -IAr, Arabic and Persian pluralization rules. We extracted tokens whose UPOS tag is NOUN with the morphological feature Number=Plur. Tokens containing -Iar or -Ier in the suffix (after the lemma) were classified as Turkish plurals. Tokens lacking these suffixes, despite being tagged as plural, were classified as Arabic or Persian plurals if they also ended with specific foreign suffixes (e.g., -āt, -ān, -īn, -hā) or contained the Gender=Fem feature, an indicator of Arabic broken plurals, annotated exclusively in the Ottoman Turkish treebank. Table 6 provides the distribution of the patterns across both corpora. In total, 224 cases out of 800 plural nouns were pluralized with either Arabic or Persian grammar rules, corresponding to 28% of all plural nouns. However, the dominant pluralization rule was still Turkish, -IAr, by 72%, which shows the significance of Turkish morphology in the Ottoman Turkish language.

Pluralization Type	Ottoman Turkish	Modern Turkish
Turkish -IAr	576	6,535
Arabic/Persian plurals	224	0

Table 6. Observed pluralization patterns

In modern Turkish, Arabic and Persian plurals have lost their function. Nevertheless, plural

words such as *eṣnā* (moment.PL) in Ottoman Turkish can be still found in TR-BOUN as *eṣna* (moment), although they are annotated as singular because their meaning does not hold any plurality anymore.

Finally, we identified five instances where a noun was first pluralized with an Arabic pluralization rule (broken plurals), and then with the Turkish pluralization suffix, *-lAr*, meaning that the word was pluralized twice. The instances of double pluralization were *evlād u nisvān u emvālleri* (their child.PL, woman.PL, and good.PL-PL-3PL.POSS), *evbāṣ-lar* (ruffian.PL-PL), and *mevācib-ler* (salary.PL-PL).

5. Discussion

This study quantitatively examines the transition from Ottoman to modern Turkish. Our quantitative findings align with the previous qualitative studies of Ottoman Turkish language. Timurtaş (1976) previously argued that Ottoman Turkish exhibited a certain tendency toward labialization, particularly in suffixes. Our data on compliance in vowel harmony confirms this claim as labial vowel harmony compliance rate in Ottoman Turkish was significantly lower than in modern Turkish. This aligns with Stachowski's (2013, 550) observation of "labially disharmonized vowel sequences" in Ottoman Turkish. While Stachowski noted these forms in grammar books, our study quantifies their prevalence in actual usage and shows that the transition to modern Turkish involved the elimination of these rounded forms in favor of the strict labial harmony described by Göksel and Kerslake (2005). Moreover, the discrepancy for the vowel harmony compliance rates can be explained by the lack of front and back vowels for some suffixes in Ottoman Turkish, like first person singular possessive suffix and genitive case suffix, as shown by Demir and Yılmaz (2002, 862). Although previous studies stated the gap in vowel harmony between Ottoman Turkish and modern Turkish, this study quantifies this gap. Thus, the study both reaffirms previous studies and demonstrates the magnitude of this gap at this scale.

Additionally, Table 4 demonstrates that palatal vowel harmony was well-established in both Ottoman Turkish and modern Turkish. A close analysis uncovers that the disharmonic instances are mainly loanwords such as *ḥakikatde* (truth-LOC) in Ottoman Turkish and *ḥakikate* (truth-DAT) in modern Turkish. The presence of persistent disharmony aligns with Göksel and Kerslake (2005) and Khalilzadeh (2010, 149). We frame these not as modern anomalies, but as remnants of Ottoman phonological constraints that resisted the general tendency in the language towards a more solid vowel harmony.

The presence of nouns that were first pluralized with Arabic pluralization rule, and then the Turkish

pluralization suffix, *-lAr* provides quantitative evidence for the mechanism of reanalysis. Following Harris and Campbell (1995, 61-119), these Arabic plurals were reanalyzed by Ottoman Turkish authors as singular stems. Consequently, the mechanism of extension was triggered to align these forms with the productive agglutinative grammar.

6. Conclusion

To sum up, this study computationally compares Ottoman and modern Turkish to quantify the morphophonological transition. By utilizing annotated treebanks in UD, we demonstrated the diachronic shift in language use, particularly for vowel harmony in noun suffixes as well as the verb suffixes, the converb *-(y)lp* and the dative infinitive *-mAyA*, and pluralization. The method goes beyond descriptive statistics and applies log-likelihood ratio test to quantify the shift further. The statistical analysis revealed that while modern Turkish achieved 98% compliance with labial vowel harmony, Ottoman Turkish complied with it by 77%. Additionally, it exhibits the obsolescence of certain allomorphs in modern Turkish. Furthermore, the analysis of plural nouns both demonstrated the disappearance of the function of Arabic and Persian pluralization rules in modern Turkish and the presence of double pluralization in Ottoman Turkish through Turkish and Arabic grammar. Future work will expand this analysis in a larger and more granular dataset in terms of period, genre, and author gender, and deploy logistic regression analysis to find the features such as period, genre, or surrounding phonological environment, that predict the observed patterns.

7. Limitations

A primary challenge in this study appears in interpreting the Ottoman Turkish data. Since Ottoman Turkish was used across six centuries, the changes in it should be acknowledged. However, due to the lack of metadata on when a given sentence was written in UD for Ottoman Turkish, the study diachronically flattens Ottoman Turkish and treats Ottoman Turkish as a monolith. Furthermore, the methodology for data collection and genres of the treebanks vary, which may affect how well they represent the language. For instance, although *-(y)lBAn* does not appear in Ottoman Turkish treebank, this does not mean the form was absent in the language at all but points to an important limitation of the data on representativeness. A simple search on canonical Ottoman Turkish sources suggests that the suffix was present in Ottoman Turkish across centuries such as *oluban* (be-CVB) in *Adlî Dîvânî* from 16th century (Sultân Bâyezîd-i Sâni, 2018, 20) and *açuban* (open-CVB) in *Mecmûa-ı Gazeliyyât* from 17th century (Konevî, 2017, 191). Thus, such

limitations should be considered while interpreting the data.

8. Bibliographical References

- Alpay, N. (2015). *Türkçe Sorunları Kılavuzu*. Metis Yayınları, 5th edition.
- Birnbaum, Eleazar. (1967). The Transliteration of Ottoman Turkish for Library and General Purposes. *Journal of the American Oriental Society*, 87(2):122–156.
- Clements, G. and Sezer, E. (1982). Vowel and Consonant Disharmony in Turkish. In H. van der Hulst & N. Smith (Eds.), *The Structure of Phonological Representations (Part II)*. Berlin, Boston: De Gruyter, pp. 213–256. <https://doi.org/10.1515/9783112423325-007>
- Cristaldi, M. P. E. (2025). Between Linguistic Standardization and Nationalism: Descriptions and Grammars of the Turkish Language from the Late Ottoman to the Early Turkish Republican Era. In É. Aussant, & F. Simon (Eds.), *Documenter Et décrire Les Langues d'Asie : Histoire Et épistémologie*. Paris: Société d'histoire et d'épistémologie des sciences du langage, pp. 241–266. <https://doi.org/10.4000/138n7>
- Demir, N. and Yılmaz, E. (2002). Ottoman Turkish. *The Turks*, 3: 853–67.
- Develi, H. (1995). *Evliya Çelebi Seyahatnâmesine göre 17. yüzyıl Osmanlı Türkçesinde ses benzeşmeleri ve uyumlar* [Sound assimilations and harmonies in 17th century Ottoman Turkish according to Evliya Çelebi's Seyahatnâme]. Türk Dil Kurumu.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):51–74.
- Eckhoff, H. M., Luraghi, S., and Passarotti, M. (Eds.). (2020). *Diachronic treebanks for historical linguistics*. John Benjamins Publishing Company. <https://doi.org/10.1075/bct.113>
- Gümüškılıç, M. (1997). *18. yüzyıl Türkçesinde ses hadiseleri* [Sound events in 18th century Turkish] [Unpublished doctoral dissertation]. Istanbul University, Faculty of Letters.
- Harris, A.C. and Campbell, L. (1995). *Historical Syntax in Cross-Linguistic Perspective*. Cambridge University Press.
- Kerslake, C. (2015). Ottoman Turkish. In L. Johanson & É. Á. Csató (Eds.), *The Turkic Languages* (2nd ed., pp. 179–202). Routledge. <https://doi.org/10.4324/9781003243809>
- Khalizadeh, A. (2010). Vowel Harmony in Turkish. *Karadeniz Araştırmaları*, 6(24):141–150.
- Konevi, N. b. M. (2017). *Mecmûa-i gazeliyyât* (M. Özdemir, Ed.). T.C. Kültür ve Turizm Bakanlığı, Istanbul.
- Lewis, G. L. (1977). *Turkish* (14th ed.). New York: David McKay Company Inc.
- Lijffijt, J. (2013). *Computational methods for comparison and exploration of event sequences* (Aalto University publication series Doctoral Dissertations 205/2013) [Doctoral dissertation, Aalto University]. <http://urn.fi/URN:ISBN:978-952-60-5475-9>
- Lijffijt, J., Säily, T., and Nevalainen, T. (2012). CEECing the baseline: lexical stability and significant change in a historical corpus. In *Variation, Contacts and Change in English*, 10. https://varieng.helsinki.fi/series/volumes/10/lijffi_jt_saily_nevalainen/ (accessed 25 December 2025)
- Göksel, A. and Kerslake, C. (2005). *Turkish: A comprehensive grammar*. Routledge.
- Montes, T., Manrique-Gómez, L., and Manrique, R. (2024). Historical Ink: Semantic Shift Detection for 19th Century Spanish. In *Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change*, pages 29–41, Bangkok, Thailand, August. Association for Computational Linguistics.
- Nevalainen, T., Raumolin-Brunberg, H., and Mannila, H. (2011). The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change* 23(1):1–43.
- Özçelik, Ö. (2024). Vowel harmony. In *The Phonology of Turkish*. Oxford: Oxford Academic.
- Petrova, Y. A. and Kuleshova, K. A. (2020). Cultural and historical evolution of the Turkish language under the influence and borrowing of European languages. *Nauchnyj al'manakh stran Prichernomor'ya [Science Almanac of Black Sea Region Countries]*, 21(1):65–70. <https://doi.org/10.23947/2414-1143-2020-21-1-65-70>
- Rayson, P., Damon B., and Brian F. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. In *Le poids des mots: Proceedings of the 7th International Conference on the Statistical Analysis of Textual Data (JADT 2004)*, pages 926–36.
- Sultân Bâyezîd-i Sâni. (2018). *Adli divânı (Divân-ı Sultân Bâyezîd-i Sâni)* (Y. Bayram, Ed.). T.C. Kültür ve Turizm Bakanlığı, Istanbul.
- Stachowski, M. (2013). On the evolution of Ottoman Turkish vowel harmony and the (il)labiality of the suffix +lık. In B. Gül, F. Ağca & F. Gökçe (Eds.), *Bengü Bitig*. Ankara: Öncü Kitap, pp. 551–558.
- Timurtaş, F. (1976). Küçük Eski Anadolu Türkçesi Grameri. *Türkiyat Mecmuası*, 18, pp. 331–368.
- Yağmur, Ö. (2025). Ottoman to modern Turkish: Persian izafet and its colloquial variations. *TEKE (Türkçe Eğitimi ve Kültürü Dergisi)*, 14(3):821–45. <https://doi.org/10.7884/teke.1639691>

9. Language Resource References

- Marşan, B., Akkurt, S. F., Şen, M., Gürbüz, M., GÜngör, O., Özateş, Ş. B., Üsküdarlı, S., Özgür, A., GÜngör, T., and Öztürk, B. (2022). *Enhancements to the BOUN Treebank reflecting the agglutinative nature of Turkish*. arXiv.
<https://doi.org/10.48550/arXiv.2207.11782>
- Yilandilođlu, E. and Siewert, J. (2025). DUDU: A treebank for Ottoman Turkish in UD style. In Š. A. Holdt, N. Ilinykh, B. Scalvini, M. Bruton, I. N. Debess, & C. M. Tudor (Eds.), *The Third Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL 2025)*, pages 74–79, Tallinn, Estonia, March. University of Tartu Library.
<https://aclanthology.org/2025.resourceful-1.16/>