

A Robust Optimization Framework for Historical Word-Sense Disambiguation in Digital Philology

Alisher Doniyorov¹, Ruzikhol Panjjeva², Nazira Normatova³, Saboxat Kabulova⁴,
Maftunakhon Azimova⁵, Zebo Musinova⁶, Dilbar Tursunova⁷

¹Tashkent State University of Oriental Studies, Uzbekistan. Email:
doniyorov_alisher@mail.ru, <https://orcid.org/0000-0001-8710-5398>

²Termez State University, Termez, Uzbekistan. Email: panjiyevarezixol@gmail.com,
<https://orcid.org/0009-0005-6479-537X>

³Researcher, Department of Methodology of Teaching the English Language, Faculty of
Language and Translation, Namangan State Institute of Foreign Languages, Namangan,
Uzbekistan.

Email: naziranormatova82@gmail.com, <https://orcid.org/0009-0009-6730-8441>

⁴Mamun University, Khiva, Uzbekistan. Email: kabulova_saboxat@mamunedu.uz,
<https://orcid.org/0009-0000-1241-1446>

⁵Senior Lecturer, Department of Social Sciences Kokand University, Kokand, Fergana
Region, Uzbekistan. Email: maftunaazimova67@gmail.com, <https://orcid.org/0009-0000-8123-9862>

⁶Department of Russian Philology, Samarkand State Institute of Foreign Languages,
Samarkand, Uzbekistan. Email: zebomusinova08@gmail.com, <https://orcid.org/0009-0009-7656-8691>

⁷Department of Pedagogy, Faculty of Pedagogy-Psychology and Inclusive Education,
Jizzakh State Pedagogical University, Email: dilbart718@gmail.com, <https://orcid.org/0009-0003-4347-0846>

Abstract

In historical word-sense disambiguation, interpretation of a language with meanings, spelling norms, and texts that might differ from those used by current language models is required. Incomplete documents, unknown dates, and insufficient dictionaries could render an unambiguous interpretation with one top score doubtful. This article discusses the use of a robust optimization method for the ranking of historical senses based on contextual, lexical, and historical evidence. Each score of evidence is accompanied by a finite adversarial deviation, and a robustness budget regulates to what extent the scores could suffer deterioration at the same time. The optimal interpretation is defined by maximizing the minimum value of its score within the given uncertainty set. The example case is concerned with three possible interpretations of the word “art.” In the nominal scoring procedure, the interpretation that is deemed most appropriate is the aesthetic one with the score 0.8650, followed by the practical skill interpretation with the score 0.8200. Under the robustness budget of one, the winning candidate would be the practical skill with the score 0.7700 and the aesthetic interpretation with the score 0.6650. However, both interpretations would still be eligible for consideration due to overlapping evidence intervals. The numeric approach shows a sensitivity to uncertainty of evidence but not the accuracy evaluated against a historical corpus. This methodology makes it possible to provide inspectable interpretation by documenting the evidence, assumptions, and perturbations involved in each ranking.

Keywords: Digital philology, historical semantics, word-sense disambiguation, robust optimization, textual uncertainty.

1. Introduction

Digital philology uses computer analysis along with detailed study of witnesses to texts, their historical background, and editorial interpretation. In a digital word, there can be elements of spelling that were used before; the word can have a different meaning compared to its modern usage, and there may be errors in the rest of the text. The use of digital scholarly edition entails the necessity to make deliberate choices regarding the ways sources and interpretations are presented (Pierazzo, 2016). For historical natural language processing, the problems are analogous (Piotrowski, 2012).

Disambiguation of word sense is commonly concerned with which sense a given word represents in a certain context (Navigli, 2009). With regard to historical texts, however, this problem requires further considerations, such as time, genre, geographical location, and the record kept by the witness. For instance, a contemporary semantic tool may give a high rank to a conventional sense of a word despite the fact that a rare one is more suitable in light of period-specific sources.

This cannot be achieved by taking the position that words that are older have only their older meanings. Words may have multiple meanings, dates are unreliable, and the later versions can contain the earlier vocabulary. There should be critical analysis of the historical information and not its replacement with a strict date criterion. It would be helpful if there were criteria to tell apart the preferred translation from a good one.

In this paper, we build a ranking approach for doing so. We make use of multiple channels of evidence, discount the scores in an uncertain set bound and retain alternatives in the case of intersecting intervals of scores. Our contribution is a straightforward application of robust optimization for token-level historical interpretation. This paper is theoretical and has an example constructed to be reproducible. There is no trained model or annotated dataset presented.

2. Historical semantics and textual evidence

2.1. Meaning change across periods

Diachronic representations help detect variations in lexical usage. Hamilton et al. studied semantic changes by using aligned historical word embeddings (Hamilton et al., 2016). Kutuzov et al. reviewed the methods and issues related to diachronic embeddings and semantic shift (Kutuzov et al., 2018). The above-mentioned tools are useful to find variations in the distributional pattern, but the variation in the representation of a word within its corpus does not determine the meaning of an instance of a word.

Frermann and Lapata described diachronic meanings as senses that become prevalent with time (Frermann & Lapata, 2016). The sense-based view of the phenomenon is important in developing period-sensitive candidate inventories. However, the frequency of the meaning of a word in a period is just contextual information but not conclusive proof that the word has a specific meaning at an instance. Innovative meanings are possible. Modeling language acquisition in multilingual settings reveals the impact of cognitive and social factors on language use Zaidi, (2025).

2.2. OCR noise and historical spelling

Historical texts may have real spelling variations along with errors introduced through digitization. The ways of processing historical spelling and textual resource preparation are described by Piotrowski, (2012). Hill and Hengchen study the influence of the quality of optical character recognition on analyses of historical texts (Hill & Hengchen, 2019). Hämäläinen and Hengchen propose a strategy for the post-processing of OCR Hämäläinen & Hengchen, (2019). All of this research supports the need to trace uncertainty in the textual data and in the semantic model. Idiomatic expression transfer analysis across contemporary novels is related to contextual meaning, cultural semantics, and ambiguity in literary texts (Avlayev et al., 2026).

Normalization is useful for retrieving candidate readings, but in case when the normalized form substitutes the witness reading, it will remove some information. The workflow proposed stores both readings and the way of conversion of one reading to another. A suggested OCR correction is considered as an alternative reading until it is confirmed by the source or any other evidence.

2.3. Token interpretation and diachronic benchmarks

The SemEval-2020 Task 1 provides an evaluation methodology for unsupervised change detection of lexical semantics (Schlechtweg et al., 2020). The resource DWUG contains diachronic word-usage graphs using contextual meaning assessments (Schlechtweg et al., 2021). While these two resources may help in historical semantical studies, change detection and sense assignment tasks are distinct from each other.

The unified WSD evaluation methodology proposed by Raganato et al. demonstrates the significance of having comparable annotations and scoring methods (Raganato et al., 2017). The historical application, however, needs an extra component – the candidate meanings and evaluation should be relevant to the period and sources. Thus, this paper presents an evaluation approach instead of stating that the performance on a semantic-change benchmark is equivalent to that on historical disambiguation. Multimodal digital information systems provide linguistic competencies and facilitate the digital language analysis in order to interpret linguistic information in digital philology (Kholodniak, 2025).

Contextual semantic analysis and textual interpretations in the area of arts and humanities with the help of large language models aid in disambiguating word senses in digital philology (Brandt & Scholz, 2024).

3. Source preparation and evidence representation

3.1. Keeping the witness and the normalized text together

Each target example is connected with an annotation consisting of the original token, normalized variants, contextual information, document ID, position in the witness, and dating information if available. Coordinates of the image source should be preserved if the source is scanned. Date annotation separates composition date, copying date, publishing date, and date of the surviving witness, if such dates are known.

This approach allows avoiding the confusion of publishing date information with the age of the language in use in the text. Transcription confidence, undeciphered tokens, and correction procedure are also preserved in this workflow. This information defines the need for widening the uncertainty margin of certain pieces of evidence.

3.2. Building a period-sensitive sense inventory

Candidate senses are formed from historical dictionaries, uses with annotation, and pertinent scholarship. It is necessary to include all the senses likely for the target rather than just the typical output of a contemporary machine. A sense should not be discounted just because its earliest attested use is later than the estimated dating of the source; dictionaries provide evidence of attested uses, not an absolute timeline for when the sense is available.

The senses list needs to maintain an unknown or undocumented sense if there is any uncertainty about the coverage. Closely associated senses may be grouped if the editorial process requires a broader distinction between senses, but the decision to do so needs to be recorded. The model cannot resurrect a sense which has been eliminated in the candidate phase.

3.3. Three complementary evidence channels

Define for every candidate s , three normalized supporting scores: contextual compatibility, lexical support, and period/genre compatibility. Contextual compatibility is the score that shows how well the candidate fits the context. Lexical support is the summary of the relevant definitions and attestations. The third score relates to the connection between the candidate and the historical context.

Let e_{si} be the score for candidate s on channel i . The nominal combined score is

$$Q_s = \sum_{i=1}^m w_i e_{si}, \quad w_i \geq 0, \quad \sum_{i=1}^m w_i = 1$$

The illustration uses $m = 3$. Scores lie between zero and one, but they are supporting evidence values, not probabilities of correctness. Weights are fixed for the example. In an empirical application, scales and weights must be validated so that one channel does not dominate merely because its numerical range is larger.

Both the definition from a dictionary and the embedding trained from it are not independent evidence sources. Their common source needs to be noted, and the correlated evidence should not carry too much weight. Weighted sum is a simple but explicit assumption, not an indication that philological evidence can be reduced to three numbers.

4. Robust interpretation model

4.1. Bounded deterioration of evidence

Let δ_{si} be the maximum amount by which a supporting score may be overstated. The uncertain score is

$$\tilde{e}_{si} = e_{si} - \delta_{si} z_i, \quad 0 \leq z_i \leq 1$$

The illustration requires $0 \leq \delta_{si} \leq e_{si}$ so that deteriorated evidence remains nonnegative. This is a one-dimensional uncertain problem: it poses the question about whether there is enough support despite the possibility of unfavorable corrections. This model does not cover all possibilities of change to the source, even those for improvement. Two-dimensional modeling would be more appropriate in the case of increases and decreases.

The admissible perturbations are restricted by a budget:

$$\mathcal{U}(\Gamma) = \left\{ z: 0 \leq z_i \leq 1, \sum_{i=1}^m z_i \leq \Gamma \right\}, \quad 0 \leq \Gamma \leq m$$

At $\Gamma = 0$, the nominal scores are used. At $\Gamma = m$, every channel can deteriorate fully. Intermediate values allow a limited total amount of adverse deviation. One budget does not imply that exactly one channel fails; it can allow for partial failure of multiple channels. It is consistent with the budgeted uncertainty model considered by Bertsimas & Sim, (2004).

4.2. Selecting the strongest lower evidence score

The robust score for candidate s is its minimum supporting value within the set:

$$R_s(\Gamma) = \min_{z \in \mathcal{U}(\Gamma)} \sum_{i=1}^m w_i (e_{si} - \delta_{si} z_i)$$

The preferred candidate is

$$s^*(\Gamma) \in \operatorname{argmax}_{s \in S} R_s(\Gamma)$$

Equivalently, the robust score subtracts the greatest admissible evidence penalty from the nominal score:

$$R_s(\Gamma) = Q_s - \max_{z \in \mathcal{U}(\Gamma)} \sum_{i=1}^m w_i \delta_{si} z_i$$

Each inner problem is a linear program. With only a budget and box bounds, it can be solved by ordering the weighted deviations $w_i \delta_{si}$ and spending the perturbation budget on the largest values first. Robust optimization approaches constitute the mathematical basis for such construction (Bertsimas et al., 2011; Vandenberghe & Boyd, 2004). The choice is candidate-wise maximin and not proof that the chosen candidate is right in all possible scenarios.

4.3. Retaining overlapping readings

A good ranking could vary even in the presence of significant uncertainty. To prevent transforming a ranking into a baseless definitive annotation, one must assign an evidence range to each candidate:

$$I_s(\Gamma) = [R_s(\Gamma), Q_s]$$

Upper boundary value is merely nominal in that the uncertainty model can only reduce confidence. A separation criterion on the score is

$$R_{s^*}(\Gamma) > \max_{r \neq s^*} Q_r$$

If this condition fails, retain all candidates whose nominal score reaches the best robust lower score:

$$S_{\text{retain}} = \left\{ s: Q_s \geq \max_r R_r(\Gamma) \right\}$$

The rule provides a conservative list of candidates. It is neither a confidence set nor does it ensure that each remaining candidate is capable of being optimal under the same perturbation. It also ensures that no candidate is thrown out just because it is not separable by the marginal bounds. The editor gets a preferred reading, remaining alternatives, and explanation of the issue.

4.4. Choosing and documenting the uncertainty settings

Uncertainty radius values should be based on the weakness of the evidential base. Context scores can be verified for various reasonable corrections to OCR or other possible context windows. Lexical scores can be re-examined for different dictionaries. Period scores can be re-tested for reasonable dating and genre assignments. Such variations could help define the boundaries, but they do not represent a statistical interval per se.

The budget must be chosen using held-out documents and disclosed as part of the interpretation process. The use of budget based on inspection of the test errors may lead to overly positive evaluations. This is because the model downplays supportive evidence and too much uncertainty hides important distinctions, while too little uncertainty keeps false preferences alive.

5. Constructed Philological Analysis

5.1. A hypothetical occurrence of the word art

This worked example depicts a hypothetical excerpt of a damaged manual that was tentatively dated to the eighteenth century. The excerpt is not a quote from an archival text and does not draw on any particular document or historical dictionary entry as its source. We look at three general editorial definitions of "art": skill/craft, aesthetic work, and deceptive skill. These definitions were selected in order to use in our mathematical example rather than for a historically exhaustive categorization.

Let us assume constant channel weights of 0.50 for context, 0.30 for lexical evidence, and 0.20 for period and genre compatibility. See table 1 for the values of our hypothetical evidence scores. Modern contextual evidence prefers aesthetic work, while period-and-genre evidence prefers practical craft.

Table 1: Assumed supporting evidence for three candidate interpretations.

Candidate interpretation	Context	Lexical support	Period and genre	Nominal score
Practical craft or skill	0.80	0.80	0.90	0.8200
Aesthetic production	0.95	0.90	0.60	0.8650
Deceptive skill	0.50	0.40	0.40	0.4500

Nominal scoring determines aesthetic formation. In the example, the potential decay is higher for its contextual and lexical score, indicating an uncertain transfer from the modern model and an uncertain dictionary mapping. Table 2 provides the assumed radii and the consistent findings at $\Gamma = 1$.

Table 2: Evidence uncertainty and robust scores with a budget of one.

Candidate interpretation	Context radius	Lexical radius	Period radius	Maximum penalty	Robust score
Practical craft or skill	0.10	0.05	0.05	0.0500	0.7700
Aesthetic production	0.40	0.25	0.05	0.2000	0.6650
Deceptive skill	0.10	0.10	0.05	0.0500	0.4000

5.2. Why the preferred interpretation changes

Practical Craft: The weighted deviations are 0.0500, 0.0150, and 0.0100. For a budget of one, the whole deviation would be assigned to the contextual channel, and thus the penalty would be 0.0500. Therefore, the robust score would be $0.8200 - 0.0500$, which equals 0.7700.

Aesthetic Production: The weighted deviations are 0.2000, 0.0750, and 0.0100. The maximum possible penalty is 0.2000, and the robust score would be 0.6650. Practical craft interpretation should be chosen because it has smaller exposure to the presumed deviations than aesthetic production.

The difference between the two robust scores is 0.1050. This is not an indication of a probability difference or higher accuracy but rather the difference in worst-case scores. The reason for changing the ranking order comes from the values provided, and it could have been different had the historical data been different.

5.3. Sensitivity to the robustness budget

Table 3 illustrates the ranking changes with the increasing budget. The computations were verified through linear programming with similar box constraints and perturbation budgets.

Table 3: Robust supporting scores under increasing uncertainty budgets.

Budget	Practical craft	Aesthetic production	Deceptive skill	Preferred reading
0.00	0.8200	0.8650	0.4500	Aesthetic production
0.25	0.8075	0.8150	0.4375	Aesthetic production
0.50	0.7950	0.7650	0.4250	Practical craft
1.00	0.7700	0.6650	0.4000	Practical craft
2.00	0.7550	0.5900	0.3700	Practical craft
3.00	0.7450	0.5800	0.3600	Practical craft

For budgets between zero and one, the practical and aesthetic robust scores are $0.8200 - 0.0500\Gamma$ and $0.8650 - 0.2000\Gamma$, respectively. They coincide at $\Gamma = 0.30$. Beyond this point, practical art has the superior robust rating within this range. The point marks the model that is constructed and not any historical demarcation of what constitutes "art."

5.4. An editorial recommendation rather than automatic closure

For a budget of one, the interval for practical-craft is $[0.7700, 0.8200]$ while that of aesthetics is $[0.6650, 0.8650]$.

The lower practical score does not exceed the upper aesthetic score. Score separation condition is thus not met and both scores will stay. The skillfulness of deception is ruled out from further consideration since its score is below 0.7700, which is the best lower score.

This output is as follows: Practical Craft is favored in this case under uncertainty considerations; Aesthetic Production cannot be resolved as an alternative; the source needs editorial clarification to determine its unique meaning. This output is more informative than stating the largest robust score as a historical fact.

6. Editorial integration and empirical evaluation

6.1. Provenance attached to each sense annotation

The record of interpretation should include the target site, reading from witnesses, alternative readings normalized, potential definitions, sources of evidence, weightages, limits of deviations, budget, selected candidate, and remaining alternatives. Any changes introduced by the editor must keep the previous record as well as mention the evidence that led to the modification.

The structured text encoding can enable the association between the annotations and the passage as well as the editorial intervention (Burnard, 2014). The suggested system should maintain its output related to semantics along with the witness text, without replacing the original text. This approach follows the idea of digital textual scholarship where computational technology plays a role in interpretation and criticism (McGann, 2016).

6.2. Validation without anachronistic leakage

Corpus studies need to develop period-sensitive annotations either by creating them anew or by reusing existing ones, involving several experts where possible. Annotations have to allow for multiple and non-resolved readings if the text allows such readings. Agreement and disagreements have to be noted rather than forcing disputed passages to have one reading only.

Training, parameter setting, and evaluation have to be done separately per document, as well as checking for duplicate authors, editions, or passages. Period-specific mappings of the dictionary and editorial remarks have to be included because they can reveal the answer through the input. Cross-period evaluations have to separate the error due to semantic changes from the error due to other changes.

DWUG can help in graded comparisons of usage and grouping of senses Schlechtweg et al., (2021); while semantic change evaluation sets give additional and complementary tests (Schlechtweg et al., 2020). An empirical study should not present change detection correlation as if it were token-level disambiguation accuracy.

6.3. Measures that reflect both resolution and uncertainty

The evaluation will provide results on the percentage of candidate recall, accuracy or macro-F1 for uniquely identified instances, the percentage of automatically resolved instances, and short-list recall for the rest. In cases where multiple senses are allowed by the annotation, the evaluation procedure should clarify how agreement between predictions is treated. Results for language, period, genre and OCR quality strata will be provided separately where the sample sizes allow.

Comparison of robustness and nominal procedures will be done at similar levels of automatic resolution. Otherwise, the method will appear more reliable merely because it refers more instances to experts. The amount of review time spent, rate of override of recommended candidates, and response to source errors are further criteria that could evaluate the usefulness of editing. Such measures need observations and cannot be generated from a constructed instance.

7. Limitations and conclusion

The model is dependent on the quality of the candidate list and its evidence channel interpretation. An absent historical meaning cannot be reconstructed through a solid ranking process. There may be correlations among scores, poorly estimated uncertainty bounds, and the model does not consider cases where increased support occurs after correction. Dates of composition, dates of witness, and dictionary attestations can also be ambiguous in a way that is not fully captured by the score.

The example represents only one hypothetical event and three manually selected interpretations. The example provides calculations and the working principle of the ranking rule, but not the performance of the process on historical data. The framework does not provide proof of correctness of the preferred interpretation, and the overlap shortlist is not a calibrated confidence set. More difficult source materials will need structured uncertainty between OCR options, document dates, and sense candidates.

In these terms, robust optimization offers a clear means of determining which historical interpretation still holds up once the vulnerable evidence is taken out of consideration. The example shifts the favored interpretation from aesthetic creation to practical artistry when the budget is higher than 0.30, but retains both interpretations for consideration when their respective ranges intersect. The key utility is thus an explicit difference between preference that is robust versus evidence that suffices for unique annotation. Future corpus studies should use estimates of uncertainty based on archival conditions and assess the method by historical expertise and resolution rates.

References

1. Avlayev, O., Lutfullaeva, D., Ziyamukhamedov, J., Yuldasheva, D., Yuldasheva, M., Suvanov, X., & Tashkenbayeva, D. (2026). Bridging philology and linguoculturology to analyze idiomatic expression transfer in contemporary novels. *Journal of Internet Services and Information Security*, 16(2), 27–39. <https://doi.org/10.58346/JISIS.2026.12.003>
2. Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations research*, 52(1), 35–53. <https://doi.org/10.1287/opre.1030.0065>
3. Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM review*, 53(3), 464–501. <https://doi.org/10.1137/080734510>
4. Brandt, F., & Scholz, K. (2024). Large language model-assisted research framework for arts and humanities knowledge discovery. *International Journal of Research in Arts and Science*, 10(1), 11–19.
5. Burnard, L. (2014). *What is the Text Encoding Initiative? How to add intelligent markup to digital resources* (No. 3). OpenEdition Press.
6. Frermann, L., & Lapata, M. (2016). A Bayesian model of diachronic meaning change. *Transactions of the Association for Computational Linguistics*, 4, 31–45. https://doi.org/10.1162/tacl_a_00081
7. Hämäläinen, M., & Hengchen, S. (2019, September). From the pafit to the fiiture: a fully automatic NMT and word embeddings method for OCR post-correction. In *Proceedings of the international conference on recent advances in natural language processing (RANLP 2019)* (pp. 431–436). https://doi.org/10.26615/978-954-452-056-4_051
8. Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016, August). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 1489–1501). <https://doi.org/10.18653/v1/P16-1141>
9. Hill, M. J., & Hengchen, S. (2019). Quantifying the impact of dirty OCR on historical text analysis: Eighteenth Century Collections Online as a case study. *Digital Scholarship in the Humanities*, 34(4), 825–843. <https://doi.org/10.1093/llc/fqz024>
10. Kholodniak, O. (2025). Developing Linguistic Competence through Multimodal Digital Information Systems – Insights from English Philology. *Indian Journal of Information Sources and Services*, 15(3), 354–361. <https://doi.org/10.51983/ijiss-2025.IJISS.15.3.40>
11. Kutuzov, A., Øvrelid, L., Szymanski, T., & Velldal, E. (2018, August). Diachronic word embeddings and semantic shifts: a survey. In *Proceedings of the 27th international conference on computational linguistics* (pp. 1384–1397).
12. McGann, J. (2016). *Radiant textuality: literary studies after the World Wide Web*. Springer. <https://doi.org/10.1007/978-1-137-10738-1>
13. Navigli, R. (2009). Word sense disambiguation: A survey. *ACM computing surveys (CSUR)*, 41(2), 1–69. <https://doi.org/10.1145/1459352.1459355>
14. Pierazzo, E. (2016). *Digital scholarly editing: Theories, models and methods*. Routledge. 1–247. <https://doi.org/10.4324/9781315577227>
15. Piotrowski, M. (2012). *Natural language processing for historical texts*. Morgan & Claypool Publishers. <https://doi.org/10.2200/S00436ED1V01Y201207HLT017>
16. Raganato, A., Camacho-Collados, J., & Navigli, R. (2017, April). Word sense disambiguation: A unified evaluation framework and empirical comparison. In *Proceedings of the 15th Conference of the European*

Chapter of the Association for Computational Linguistics: Volume 1, Long Papers (pp. 99-110). <https://doi.org/10.18653/v1/E17-1010>

17. Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., & Tahmasebi, N. (2020, December). SemEval-2020 task 1: Unsupervised lexical semantic change detection. In *Proceedings of the fourteenth workshop on semantic evaluation* (pp. 1-23). <https://doi.org/10.18653/v1/2020.semeval-1.1>

18. Schlechtweg, D., Tahmasebi, N., Hengchen, S., Dubossarsky, H., & McGillivray, B. (2021, November). DWUG: A large resource of diachronic word usage graphs in four languages. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 7079-7091). <https://doi.org/10.18653/v1/2021.emnlp-main.567>

19. Vandenberghe, L., & Boyd, S. (2004). *Convex optimization* (Vol. 1). Cambridge: Cambridge University press. <https://doi.org/10.1017/CBO9780511804441>

20. Zaidi, Z. F. (2025). Developing a Predictive Model for Language Acquisition in Multilingual Environments Based on Cognitive and Social Factors. *International Academic Journal of Humanities*, 12(3), 1-8. <https://doi.org/10.71086/IAJH/V12I3/IAJH1208>

21. Mukti, I. Z., Khan, E. R., Biswas, K. K. (2026). Affordance-Learning Middleware For Zero-Configuration Interaction & With Everyday Objects. *National Journal of Ubiquitous Computing and Intelligent Environments*, 40-46.