A Pilot Computational Test for Complexity Redistribution: Latin→Early Spanish Case Study

Abstract

We present a pilot computational method to test whether grammatical complexity redistributes across subsystems during language change. We analyze 16 texts spanning Classical/Medieval Latin and Early Spanish (n=7/5/4) with Stanza-based pipelines and non-parametric statistics. Two metrics vary strongly by period—Spanish article rates rise from 0 to approximately 107 per 1,000 words and analytic auxiliaries replace Latin synthetic forms—while dependency-depth differences are not statistically significant (Kruskal–Wallis p=0.114). A known modeling limitation under-detects Medieval Latin analytic forms, creating a discontinuity that precludes full evolutionary claims. We position our results as method validation rather than hypothesis confirmation, and we outline requirements for a definitive study (specialized Medieval models; >=20 texts/period; genre controls). Code and data are provided for replication.

Keywords: historical linguistics, complexity redistribution, computational linguistics, Latin, Spanish, grammatical evolution

1. Introduction

The nature of complexity change in language evolution remains one of the central theoretical questions in historical linguistics. Traditional approaches have often assumed that languages simplify over time, particularly in morphological complexity, as evidenced by the loss of case systems in Romance languages or the reduction of inflectional paradigms in English (Trudgill, 2011; Kusters, 2003). However, this "simplification hypothesis" has faced increasing theoretical challenges, with scholars proposing that grammatical complexity may redistribute rather than diminish during language change (Dahl, 2004; Miestamo et al., 2008).

The complexity conservation hypothesis suggests that languages maintain a relatively constant "complexity budget" across different grammatical subsystems. When complexity is lost in one area (e.g., morphological case marking), it is compensated by increased complexity elsewhere (e.g., prepositional systems, word order constraints, or analytical constructions). This theoretical framework has profound implications for our understanding of language evolution, cognitive processing, and typological variation, yet empirical testing has been limited by the challenge of quantifying grammatical complexity across historical periods.

Recent advances in computational linguistics and natural language processing provide new opportunities to test complexity conservation hypotheses through large-scale quantitative analysis of historical texts. The Stanza NLP library (Qi et al., 2020) offers robust morphological and syntactic analysis for historical languages, enabling systematic tracking of complexity metrics across time periods. Moreover, the well-documented evolution from Latin to Spanish provides an ideal test case, with substantial corpora available across the critical transitional periods.

This study addresses three key research questions: (1) Does Spanish article development compensate for Latin case marking loss? (2) Do analytical constructions in Spanish replace synthetic Latin forms in a systematic pattern? (3) Do dependency complexity patterns show compensatory mechanisms during language transition?

2. Theoretical Framework

2.1 Complexity Conservation Theory

The complexity conservation hypothesis, rooted in functionalist and usage-based approaches to language change, proposes that cognitive and communicative pressures maintain relatively stable overall complexity levels across linguistic systems (Dahl, 2004; Hawkins, 2009). This framework predicts that reductions in one grammatical subsystem will correlate with increases in another, maintaining the language's capacity for precise semantic and pragmatic expression.

Key mechanisms include: (1) **Compensatory development**: New grammatical elements emerge to replace lost functions (e.g., articles replacing case markings); (2) **Redistribution**: Complexity shifts between subsystems (e.g., from morphology to syntax); (3) **Functional**

preservation: Core communicative functions are maintained through alternative structural means.

2.2 Latin to Spanish Evolution

The evolution from Latin to Spanish provides exceptional evidence for complexity conservation mechanisms. Classical Latin employed a six-case system (nominative, accusative, genitive, dative, ablative, vocative) with extensive synthetic verbal morphology. Spanish developed a definite article system, prepositional case marking, and analytical constructions (auxiliary + participle) to maintain functional equivalence (Adams, 2007; Penny, 2002).

Previous descriptive studies have identified these transitions qualitatively, but quantitative analysis of complexity redistribution has been lacking. Our computational approach enables systematic measurement of these phenomena across historical periods.

3. Methods

3.1 Corpus Design

We analyzed 16 historical texts representing three critical periods:

Classical Latin (7 texts, ~1st century BCE - 1st century CE): - Caesar: *Commentarii de Bello Gallico* (Books 1-2) - Cicero: *Epistulae ad Familiares* (Books 1-2) - Livy: *Ab Urbe Condita* (Book 1) - Sallust: *Bellum Catilinae* (Book 1), *Bellum Iugurthinum* (Book 1)

Medieval Latin (5 texts, ~6th-8th centuries CE): - Gregory of Tours: *Decem Libri Historiarum* (Book 1) - Bede: *Historia Ecclesiastica Gentis Anglorum* (Book 1) - Isidore of Seville: *Etymologiae* (Books 1-2, selected sections) - *Peregrinatio Egeriae* (complete text) - Anonymous chronicles (selected)

Early Spanish (4 texts, ~12th-13th centuries): - *Auto de los Reyes Magos* (~1150) - *Cantar de mio Cid* (~1200) - Gonzalo de Berceo: *Milagros de Nuestra Señora* (~1260) - *Fuero de Peñafiel* (1264)

Text selection prioritized: (1) chronological representation across transitional periods; (2)

textual authenticity and scholarly consensus on dating; (3) sufficient length for statistical analysis; (4) genre diversity within constraints of available medieval sources.

3.2 Computational Analysis

NLP Processing: We employed the Stanza NLP library (version 1.5.x) with pre-trained models for Latin ('la') and Spanish ('es'). Latin texts were processed using the Latin Internet Tree Bank (ITTB) model; Spanish texts used the AnCora treebank model.

Complexity Metrics:

1. **Article Development**: Quantified definite and indefinite article usage, normalized per 1000 words. Language-specific processing distinguished actual Spanish articles ('el', 'la', 'los', 'las', 'un', 'una') from Latin determiners ('hic', 'ille', 'omnis'), which were excluded from article counts as Latin lacks an article system. Categories included definiteness marking and case-marking function.

2. **Analytical Construction Analysis**: Tracked synthetic vs. analytical forms across verb categories:

   - Future tense: Latin *amabit* vs. Spanish *va a amar*
   - Perfect aspect: Latin *amavit* vs. Spanish *ha amado*

   - Passive voice: Latin *amatur* vs. Spanish *es amado*
   - Conditional mood: Latin *amaret* vs. Spanish *amaría*

3. **Dependency Complexity**: Measured syntactic complexity through:

   - Average dependency depth per sentence
   - Maximum dependency depth per text
   - Argument structure complexity (core vs. non-core dependencies)

3.3 Statistical Analysis

Given the historical nature of the data and necessarily small sample sizes typical in historical linguistics, we employed non-parametric statistical methods:

- **Kruskal-Wallis H tests** for overall period comparisons
- **Mann-Whitney U tests** for pairwise comparisons
- **Fisher's exact tests (2×2)** for categorical data (synthetic vs. analytical constructions), specifically Classical vs. Early Spanish. A 3×2 chi-square is not appropriate due to a structural zero for Medieval Latin (model underdetection); Medieval counts are reported descriptively with caveats.
- **Bootstrap confidence intervals** (n=1000) for robust uncertainty estimation
- report medians with bootstrap BCa 95% CIs; effect sizes are not reported in this pilot due to small n and zero-variance cells

Statistical significance was set at alpha = 0.05. Multiple comparisons were controlled with Bonferroni correction (alpha = 0.017); key findings remain significant after correction. Effect sizes were not calculated for comparisons where one group showed zero variance, as this yields undefined or infinite values. All analyses were conducted in Python using SciPy (version 1.9+).

3.4 Data & Models: Known Failure Modes

Stanza's Latin model is trained primarily on Classical data and under-detects transitional Medieval analytic forms. We therefore restrict inferential tests on the analytic shift to Classical vs. Early Spanish and treat Medieval counts as descriptive only. This known limitation produces a structural zero for the Medieval row in the construction table and motivates the future use of specialized Medieval models.

4. Results

4.1 Article Development Analysis

Article development analysis revealed a dramatic structural transition across periods (Kruskal-Wallis H = 14.619, p = 0.0007), consistent with Spanish developing an article system to replace Latin case distinctions.

Period Comparisons (rates per 1000 words): - Classical Latin: mean = 0.000, SD = 0.000,

95% CI: [0.000, 0.000], n = 7 - Medieval Latin: mean = 0.000, SD = 0.000, 95% CI: [0.000, 0.000], n = 5
- Early Spanish: mean = 107.018, SD = 22.970, 95% CI: [85.774, 122.359], n = 4

Individual Spanish Text Rates: - Auto de los Reyes Magos: 73.778 per 1000 words - Cantar de mio Cid: 109.578 per 1000 words
- Berceo Milagros: 121.764 per 1000 words - Fuero de Peñafiel: 122.954 per 1000 words

The increasing article rates from earlier to later Spanish texts (73.8 to 123.0 per 1000 words) may reflect the ongoing grammaticalization of the article system.

Pairwise Analyses: - Classical vs. Medieval Latin: Mann-Whitney U = 17.500, p = 1.0000 (no difference - both lack articles) - Medieval Latin vs. Spanish: Mann-Whitney U = 0.000, p = 0.0108 (maximal difference between groups) - Classical Latin vs. Spanish: Mann-Whitney U = 0.000, p = 0.0031 (maximal difference between groups) Under Bonferroni adjustment for three pairwise tests (alpha_adj = 0.017), Classical vs Spanish (p = 0.0031; p_adj = 0.0093) and Medieval vs Spanish (p = 0.0108; p_adj = 0.0324) both remain significant.

The pattern shows complete absence of articles in both Latin periods, followed by systematic article development in Spanish. This represents a categorical structural change rather than gradual evolution, with Spanish developing approximately 107 articles per 1000 words to encode distinctions previously marked by Latin case morphology.

4.2 Analytical Construction Shift

Analysis shows a clear pilot signal of functional replacement, with synthetic Latin forms giving way to analytic auxiliaries in Early Spanish. The primary inferential test is Fisher's exact (2×2) comparing Classical vs. Early Spanish (p = 0.0001) and remains significant under Bonferroni adjustment (alpha_adj = 0.017). A 3×2 chi-square across Classical/Medieval/Spanish is not applicable due to a zero Medieval row; Medieval values are reported descriptively and interpreted with caution given model limitations.

Construction Inventories: - Classical Latin: 6,182 synthetic constructions, 0 analytical constructions - Medieval Latin: 0 constructions detected (transitional period - see limitations below) - Early Spanish: 0 synthetic constructions, 22 analytical constructions

Counts reflect only the predefined categories (future, perfect, passive, conditional). Spanish may exhibit other synthetic forms; we treat these as out-of-scope for this pilot inventory to maintain comparability.

Specific Spanish Analytical Patterns: - Perfect aspect: *he/has/ha + past participle* (14 instances) - Passive voice: *es/son + past participle* (8 instances) - Future constructions: Analytical forms not detected in Early Spanish texts (later development)

The absence of analytical constructions in Classical Latin and their presence in Early Spanish provides a clear pilot signal of functional replacement rather than simple addition of complexity.

4.3 Dependency Complexity Evolution

Dependency complexity analysis revealed compensatory patterns during transitional periods, though overall differences did not reach statistical significance (Kruskal-Wallis H = 4.343, p = 0.1140).

Period Statistics: - Classical Latin: mean = 3.973 depth, SD = 0.283, 95% CI: [3.756, 4.174], n = 7 - Medieval Latin: mean = 8.734 depth, SD = 5.266, 95% CI: [4.803, 13.619], n = 5 - Early Spanish: mean = 4.516 depth, SD = 1.070, 95% CI: [3.339, 5.591], n = 4

Pairwise Comparisons: - Classical vs. Medieval: Mann-Whitney U = 7.000, p = 0.1061, Cohen's d = -1.276 - Medieval vs. Spanish: Mann-Whitney U = 16.000, p = 0.1905, Cohen's d = 0.932 - Classical vs. Spanish: Mann-Whitney U = 7.000, p = 0.2303, Cohen's d = -0.719

Dependency depth did not differ across periods (Kruskal–Wallis p = 0.114, n=7/5/4). We therefore refrain from further interpretation and treat these results as a power baseline for a larger study.

4.4 Individual Text Analysis

Classical Latin Complexity Range: 3.59-4.46 dependency depth - Most complex: Cicero *Epistulae* Book 1 (4.46) - Least complex: Livy *Ab Urbe Condita* Book 1 (3.59)

Medieval Latin Complexity Variation: 3.42-18.38 dependency depth - Extreme outlier: Gregory of Tours *Historiarum* (18.38, max depth 98) - Most stable: Isidore *Etymologiae* (3.42)

Early Spanish Stabilization: 2.89-5.89 dependency depth - Range compression suggests systematic grammaticalization

5. Discussion

5.1 Signals Relevant to Complexity Conservation

Our findings provide pilot signals consistent with complexity conservation during Latin-to-Spanish evolution. The data show patterns consistent with this hypothesis:

Article Compensation: The dramatic article development pattern ($p = 0.0007$) correlates with case system loss. Spanish developed a systematic article system (mean 107 per 1000 words) from complete absence in Latin, potentially maintaining semantic distinctions previously encoded by case morphology through syntactic means.

Analytical Substitution: The complete transition from synthetic to analytical constructions ($p = 0.0001$) demonstrates systematic functional replacement. Spanish auxiliary constructions preserve aspectual and voice distinctions previously encoded through Latin morphological complexity.

Dependency depth: Did not differ across periods (Kruskal–Wallis $p = 0.114$, n=7/5/4). We treat this as a non-result and a power baseline for future work.

5.2 Theoretical Implications

These results are consistent with complexity conservation frameworks rather than simplification models:

Functional Maintenance: Spanish appears to develop alternative encoding strategies that maintain communicative distinctions previously marked by Latin morphological systems, though through different structural means.

Redistribution Patterns: The data suggest complexity shifts between grammatical subsystems (morphology to syntax, case marking to article systems) rather than overall reduction.

Processing Considerations: The stabilization of complexity levels in Spanish may reflect processing constraints on complexity distribution, though this requires further investigation across additional language families.

5.3 Methodological Contributions

This study demonstrates computational methods for testing complexity conservation hypotheses:

Quantitative Historical Linguistics: NLP tools enable systematic analysis of historical complexity patterns previously limited to impressionistic description.

Appropriate Statistical Methods: Non-parametric approaches accommodate the small sample sizes inherent in historical linguistic research while maintaining analytical rigor.

Cross-Period Comparisons: Standardized metrics allow direct comparison across historically distant language states.

5.4 Limitations and Future Directions

This is a pilot with very small samples per period (n=7/5/4), so all inferences are power-limited. Stanza's Latin model, trained on Classical data, under-detects transitional Medieval analytic forms; as a result, we restrict inferential tests of analytic shift to Classical vs. Early Spanish and treat Medieval counts descriptively. Genre/register imbalances may confound period contrasts; future work will add genre-matched subsets and >=20 texts per period.

Future work: - Specialized Medieval Latin tagger/lemmatizer; evaluate on held-out medieval gold. - Corpus expansion to >=20 texts/period with genre controls; preregister dependency metrics. - Add morphosyntactic complexity indices beyond depth (argument optionality, subordination rates, information-theoretic syntax surprisal).

6. Conclusions

Given the very small samples (n=7/5/4) and a Medieval modeling gap, this study provides pilot signals consistent with complexity redistribution during Latin→Spanish change. Two metrics show clear cross-period differences (articles, analytical constructions), while dependency depth does not; together these results motivate a larger, genre-controlled study with dedicated Medieval models. We release code and data to enable that next step and invite corpus contributions.

Acknowledgments

We thank the developers of the Stanza NLP library and the Perseus Digital Library for providing essential computational and textual resources. We acknowledge the limitations of automated analysis of historical texts and the ongoing need for philological expertise in computational historical linguistics.

References

Adams, J. N. (2007). *The Regional Diversification of Latin 200 BC - AD 600.* Cambridge University Press.

Dahl, Ö. (2004). *The Growth and Maintenance of Linguistic Complexity.* John Benjamins.

Hawkins, J. A. (2009). An efficiency theory of complexity and related phenomena. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language Complexity as an Evolving Variable* (pp. 252-268). Oxford University Press.

Kusters, W. (2003). *Linguistic Complexity: The Influence of Social Change on Verbal Inflection.* LOT.

Miestamo, M., Sinnemäki, K., & Karlsson, F. (Eds.). (2008). *Language Complexity: Typology, Contact, Change.* John Benjamins.

Penny, R. (2002). *A History of the Spanish Language* (2nd ed.). Cambridge University Press.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 101-108).

Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity.* Oxford University Press.

Corresponding author: Zach Nielsen (zachnielsen@hey.com)

Data availability: Code and analysis — https://github.com/nielsenz/Complexity-Conservation-in-Language-Evolution; Display bundle — https://github.com/nielsenz/Complexity-Conservation-in-Language-Evolution-Display (DOI: https://doi.org/10.5281/zenodo.17363668)
