



Retranslations Across Millennia: A Diachronic Contrastive Corpus for Studying Interlingual and Intralingual Language Contact

The evolution of valency over time: Digital pathways uncovering
DELPHI Workshop 2025, November 26-29, 2025 · Delphi, Greece

1



KEY QUESTION

What is the AthDGC project?

- What are the main research goals?
- How does it combine traditional and digital methods?
- What makes this approach innovative?

Aim of AthDGC

Creation of:

- A) A parallel diachronic corpus of influential texts
- B) A diachronic Valency Lexicon for historical languages

Analysis of:

- Valency — verb argument structure patterns
- Language diachrony — change across millennia
- Written Language Contact — translation effects

H.F.R.I. Project Number: 15284

Research Team

- **Professor Nikolaos Lavidas** (Principal Investigator)
- **Professor Emeritus Vasiliki Nikiforidou**
- **Theodoros Michalareas**
- **Vasiliki Geka**
- **Vasilis Symeonidis**
- **Sophia Chionidi**
- **Eleni Plakoutsi**
- **Anastasia Tsiropina**



Theoretical Framework

Language Contact Through Translation (LCTT)

Language Contact Through Translation (LCTT)

(Kranich, Becher & Höder 2011; McLaughlin 2011)

Key Framework:

- Translation as a venue for **language contact**
- Translations may introduce **source language features** into target languages
- Creates controlled environment for studying **contact-induced change**
- Distinguishes between **direct** and **indirect** contact effects

Unlike spoken contact situations, translation creates a textual interface where source language structures may influence target language productions. This "translationese" effect can be measured and tracked over time.

Previous Research

Digital Corpora:

- **PROIEL** (University of Oslo) — Pragmatic Resources in Old Indo-European Languages

(Haug & Jøhndal 2008; Eckhoff et al. 2018)

Synchronic Valency Lexica:

- **ValPaL** (Malchukov & Comrie 2015)

Historical/Crosslinguistic Valency Lexica:

- **HoDeL** (Zanchi 2021)
- **PaVeDa** (Luraghi et al. 2024)



The AthDGC Corpus

Influential texts across Indo-European languages

The AthDGC Corpus

Influential Texts:

Texts that have a great impact in the history of humanity (e.g., The Bible, the Homeric Poems)

Criterion: Retranslated and republished across many periods

Indo-European Languages (Phase I):

- **Hellenic:** Greek (Ancient, Koine, Medieval, Modern)
- **Germanic:** English (Old, Middle, Modern) & Gothic
- **Romance:** Latin & Old French

Original Texts

Greek Texts:

- **Ancient Greek:** Homeric poems
- **Hellenistic Greek:** New Testament
- **Medieval Greek:** Psellos' Chronographia
- **Early Modern Greek:** Digenis Akritas

English Texts:

- **Old English:** Ælfric's Lives of Saints
- **Middle English:** Orrmin: The Ormulum
- **Modern English:** Jane Austen's Pride and Prejudice

Retranslations

Intralingual Translations:

Translations **within the same language** across time periods

Example: Greek retranslation of the New Testament by Alexandros Palles (20th century)

Interlingual Translations:

Translations **between different languages**

Example: English retranslation of the New Testament by William Tyndale (16th century)

The corpus systematically includes both types to distinguish **contact-induced change** from **natural language evolution**.



PROIEL Framework Integration

Pragmatic Resources in Old Indo-European Languages

The PROIEL Platform

(Haug & Jøhndal 2008; Eckhoff et al. 2018)

Features:

- **Open-source** software tools and annotation schemes
- **Dependency grammar** scheme — effective for free-word order languages
- **Morphological and Syntactic** tagset

Annotation Guidelines:

- Tokenization and sentence division (fully automated)
- Morphological annotation (partly automated)
- Syntactic annotation (manually completed)



The Valency Lexicon

Tracking argument structure across millennia

What is Valency?

(Tesnière 1959)

Valency describes the **presence/absence, type and number of arguments** licensed by a verb:

- *Mary sleeps.* — 1 argument (monovalent)
- *Mary likes John.* — 2 arguments (divalent)
- *Mary gave a letter to John.* — 3 arguments (trivalent)

Underlying Idea:

The **verb** is the central structural element of the sentence; all other elements depend on it (dependency grammar).

Number of participants in **semantics** = number of arguments in **syntax**

Valency Alternations

(Levin 1993: up to 50 alternations for English)

Middle Alternation:

Jane broke the crystal. → *The crystal broke.*

Causative/Inchoative Alternation:

They stood the statue on the pedestal. → *The statue stood on the pedestal.*

Dative Alternation:

They gave him cake. → *They gave cake to him.*

Unspecified Object Alternation:

Mike ate all the cake. → *Mike ate.* ✓

Lexicon Example: Modern English 'write'

Modern English — Frequency: 1015

- "*You write uncommonly fast.*"
- "...*because he does not write with ease.*"
- "...*a person who can write a long letter with ease...*"
- "...*and when I next write to her...*"
- "...*it had been written five days ago.*"

Frames: NOM.V / NOM.V.ACC / NOM.V.to-DAT / Passive

Lexicon Example: Old English 'writan'

Old English — Frequency: 10

- "*Moyses lyfde þæt man write hiwgedales boc & hi forlete.*"
(Moses allowed the man to write a bill of divorce and abandon her.)
- "*for eower heortan heardnesse he eow wrat þis bebod.*"
(for your heart's hardness he wrote you this command)

Note: Dative recipient (eow = you.DAT), no preposition needed

Change observed: Loss of dative case → preposition required in Modern English



Challenges Encountered

Technical and linguistic obstacles

Challenge: Orthographic Variation

Old and Middle English Examples:

- *u* instead of *v*, *i* instead of *j*
- Same word, different spellings: *gyue*, *yyue*, *yiue*, *giue* = "give"

Our Decision:

Keep original orthography BUT create **custom dictionaries** for each text

Consequences:

- Automatic dictionary does not recognize different tokens of same word
- Harder tokenization and morphological parsing

Automatic Parsing: Late Medieval Greek

Evaluation:

Tested 4 parsers trained on Modern or Ancient Greek against reference text (Sphrantzes)

Results:

- Very low F1 scores (**<0.42**) — poor performance on Late Medieval Greek
- Better results on shorter sentences
- PROIEL has more fine-grained distinctions than UD-CoNLL

Conclusion:

Semi-automatic annotation currently as time-consuming as manual; need better conversion pipeline and trained models

English Valency Change: Parallels to Greek

(cf. Trips & Stein 2019; Trips 2020)

Similar Patterns:

- Case loss (Old English → Middle English)
- Increased preposition use
- Fixed word order development

Differences:

- **English:** Almost complete case loss
- **Greek:** Retained nominative/accusative distinction
- **English:** Stronger word order constraints

Ongoing Work

- Expand on the least represented languages (Old and Middle French, Old English)
- Add new languages (**Gothic**)
- Automate syntactic annotation:
 - A more accurate conversion pipeline from UD-CoNLL to PROIEL
 - Training existing syntactic parsers
- Collaboration with **Harvard CHS** (ARCAS Team)

Selected References (1/2)

Eckhoff, H. et al. (2018). The PROIEL treebank family. *Language Resources and Evaluation*, 52(1), 29–65.

Haug, D. & Jøhndal, M. (2008). Creating a Parallel Treebank of the Old Indo-European Bible Translations. *LREC*.

Kranich, S., Becher, V., & Höder, S. (2011). A tentative typology of translation-induced language change. In *Multilingual Discourse Production* (pp. 11–43). John Benjamins.

Lavidas, N. (2021). *The Diachrony of Written Language Contact*. Brill.

Levin, B. (1993). *English Verb Classes and Alternations*. University of Chicago Press.

McLaughlin, M. (2011). *Syntactic Borrowing in Contemporary French*. Legenda.

Selected References (2/2)

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Klincksieck.

Trips, C. & Stein, A. (2019). Contact-induced changes in argument structure of Middle English verbs. *Journal of Language Contact*, 12(1), 232–267.

Trips, C. (2020). Copying of argument structure. In *Historical Linguistics 2017* (pp. 409–430). John Benjamins.

Nikiforidou, K. (2018). Genre and constructional analysis. *Pragmatics & Cognition*, 25(3), 543–575.

Fried, M. & Nikiforidou, K. (Eds.). (2025). *The Cambridge Handbook of Construction Grammar*. Cambridge University Press.



Thank You!

Questions & Discussion

AthDGC Project · Athens Digital Glossa Chronos

nlavidas@enl.uoa.gr