

How Much of the Translator Is in a Translation?

The Targum Corpus and Measuring the Translator's Intervention

Maciej Rapacz · Aleksander Smywiński-Pohl

8 czerwca 2026

One Verse, Many Translations

KJV (1611):	Give us this day our daily bread.
The Message (2002):	Keep us alive with three square meals.
EIB (Przekład literacki, 2016):	Prosimy Ciebie, daj nam dziś naszego powszedniego chleba.
Przekład Toruński (2023):	Chleba naszego codziennego daj nam dzisiaj.
Biblia Tysiąclecia (Wyd. 5, 2003):	Naszego chleba powszedniego daj nam dzisiaj;

Examples of English and Polish translations (Mt 6:11).

What Is Translator Intervention?

“From the point of view of the target literature, all translation implies a degree of manipulation of the source text for a certain purpose.” Hermans 1985, Manipulation School

Translator intervention is any departure from a literal rendering of the source text through which the translator adapts the text to the target language, audience, or communicative purpose.

KJV (1611):	Give us this day our daily bread.
The Message (2002):	Keep us alive with three square meals.
EIB (Przekład literacki, 2016):	Prosimy Ciebie, daj nam dziś naszego powszedniego chleba.
Przekład Toruński (2023):	Chleba naszego codziennego daj nam dzisiaj.
Biblia Tysiąclecia (Wyd. 5, 2003):	Naszego chleba powszedniego daj nam dzisiaj;

Examples of the Greek original and two English translations (Mt 6:11).

How can we quantify the translator's intervention?

Agenda

1. A corpus for evaluation
2. Our proposed method

A popular aggregator of Polish Bible translations: **bibliepolskie.pl**

- around **45 translations** with full NT text in the index
- around 20 available to read online
- question: how many of them are available in existing corpora?

Stary i Nowy Testament [24]					
Biblia Leopolda	1561	Biblia warszawska [brytyjka]	1975	Śląskie Towarzystwo Biblijne	2012
Biblia Brzeska	1563	Biblia Poznańska	1975	Stare i Nowe Przymierze [EIB] - liter.	2016
Biblia Szymona Budnego	1572	Biblia Lubelska	1991	Biblia pierwszego Kościoła	2016
Biblia Wujka	1599	Interlinia - Vocatio	1993	Przekład toruński	2017
Biblia Gdańska	1632	Biblia warszawsko-praska	1997	Biblia Impulsy	2017
Przekład Mariawitów	1925	Przekład Nowego Świata	1997	Biblia Ekumeniczna	2018
Komentarze KUL	1959	Nowy Komentarz Biblijny	2005	Stare i Nowe Przymierze [EIB] - dosł.	2018
Biblia Tysiąclecia	1965	Biblia Paulistów	2008	Nowy Przekład Dynamiczny [NPD]	2021
Biblia Królowej Zofii	1455	Przekłady Izaaka Cylkowa	1883	Biblia Hebrajska	2021
Tylko Nowy Testament [21]					
NT Królewiecki	1553	NT z Wulgaty - E.Dąbrowski	1947	Ekumeniczny Przekład Przyjaciół	2012
NT Krakowski	1556	NT - Seweryn Kowalski	1957	Przekład Odzyskiwania	2018
NT Marcina Czechowica	1577	NT z greckiego - E.Dąbrowski	1961	Przekład Filologiczny	2019
NT Rakowski	1606	Słowo Życia - Parafraza NT	1989	NT - Andrzej Mazurkiewicz	2019
NT Rzewuskiego	1868	Współczesny przekład	1991	Oblubienica.eu - Przekład dosłowny	2020
NT TBS - Karol Węgierski	1876	Komentarz żydowski NT	2004	Przekład prawosławny	2022
Biblia Gdańska - Rewizja warsz.	1881	Oblubienica.eu - Przekład interlinearny	2011	Biblia Króla Jakuba	2024

Source: <https://bibliepolskie.pl/przeklady.php>

Existing Corpora vs. Translation Reality

Corpus	pol
Christodouloupoulos & Steedman (2015)	1
eBible (Akerman, 2023)	2
Mayer & Cysouw (2014)	6

Existing Corpora vs. Translation Reality

Corpus	fra	pol
Christodouloupoulos & Steedman (2015)	1	1
eBible (Akerman, 2023)	4	2
Mayer & Cysouw (2014)	17	6

Existing Corpora vs. Translation Reality

Corpus	eng	fra	ita	pol	spa	Total
Christodouloupoulos & Steedman (2015)	2	1	1	1	1	6
eBible (Akerman, 2023)	24	4	2	2	5	37
Mayer & Cysouw (2014)	39	17	7	6	22	91

Existing Corpora vs. Translation Reality

Corpus	eng	fra	ita	pol	spa	Total
Christodouloupoulos & Steedman (2015)	2	1	1	1	1	6
eBible (Akerman, 2023)	24	4	2	2	5	37
Mayer & Cysouw (2014)	39	17	7	6	22	91
Targum	390	78	33	48	102	651
Targum editions	194	41	17	29	53	334

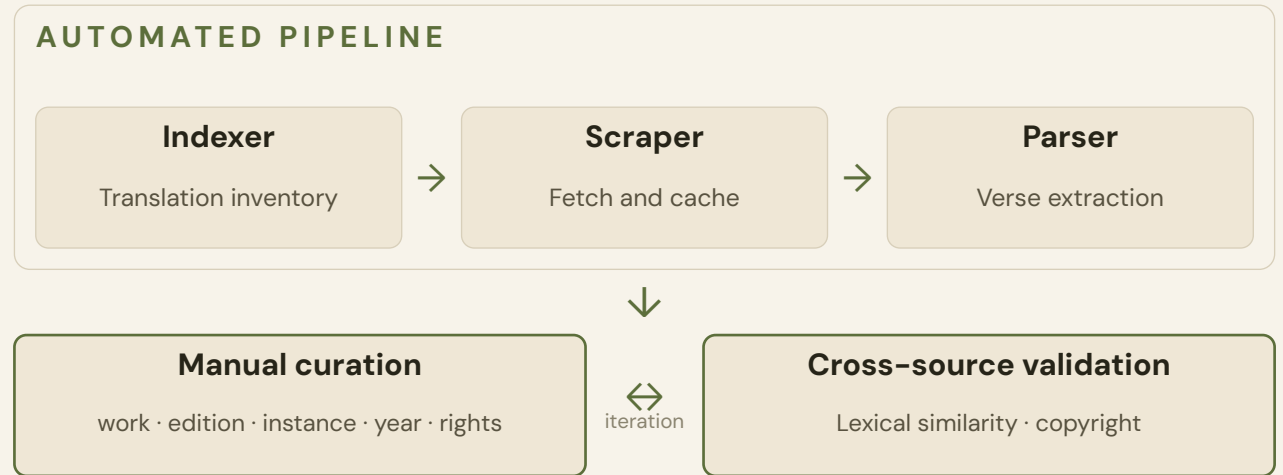
Existing Corpora vs. Translation Reality

Corpus	eng	fra	ita	pol	spa	Total
Christodouloupoulos & Steedman (2015)	2	1	1	1	1	6
eBible (Akerman, 2023)	24	4	2	2	5	37
Mayer & Cysouw (2014)	39	17	7	6	22	91
Targum	390	78	33	48	102	651
Targum editions	194	41	17	29	53	334
vs. largest	5.0×	2.4×	2.4×	4.8×	2.4×	3.7×

Translation Instances by Source

Source	eng	fra	ita	pol	spa	Total
bible.audio	3	15	–	–	1	19
bible.com	75	19	7	8	28	137
bible.is	18	8	1	2	15	44
biblegateway.com	64	4	5	3	19	95
biblehub.com	46	–	–	–	–	46
bibles.org	72	6	3	3	14	98
biblestudytools.com	37	–	–	–	–	37
bibliepolskie.pl	–	–	–	20	–	20
crossbible.com	12	1	–	–	1	14
ebible.org	24	4	2	2	5	37
jw.org	1	1	1	1	1	5
laparola.net	9	3	11	–	2	25
obohu.cz	29	17	3	9	16	74
Total	390	78	33	48	102	651

- First **scrape & cache**, then parse
- Every text is linked to a **work**, an **edition**, and a site **instance**
- The validation loop catches parser errors and duplicates



Manual Metadata Curation

Work

the translation tradition that holds editions together

Edition

a specific published revision of a work

Instance

one site-specific copy that points to an edition

Year

publication year; position on the timeline

```
work_id: king-james-version
display_name: King James Version

# Edition-level metadata
editions:
  - edition_id: king-james-version-1769
    publication_year: 1769
    copyright_status: public_domain
    copyright_reason: verified public-domain edition
    copyright_statement_checked_at: 2026-05-24

# Map site-specific instance to concrete edition
instances:
  - site: bible.com
    instance_id: KJV
    edition_id: king-james-version-1769
    edition_version_reasoning: site label matches KJV
```

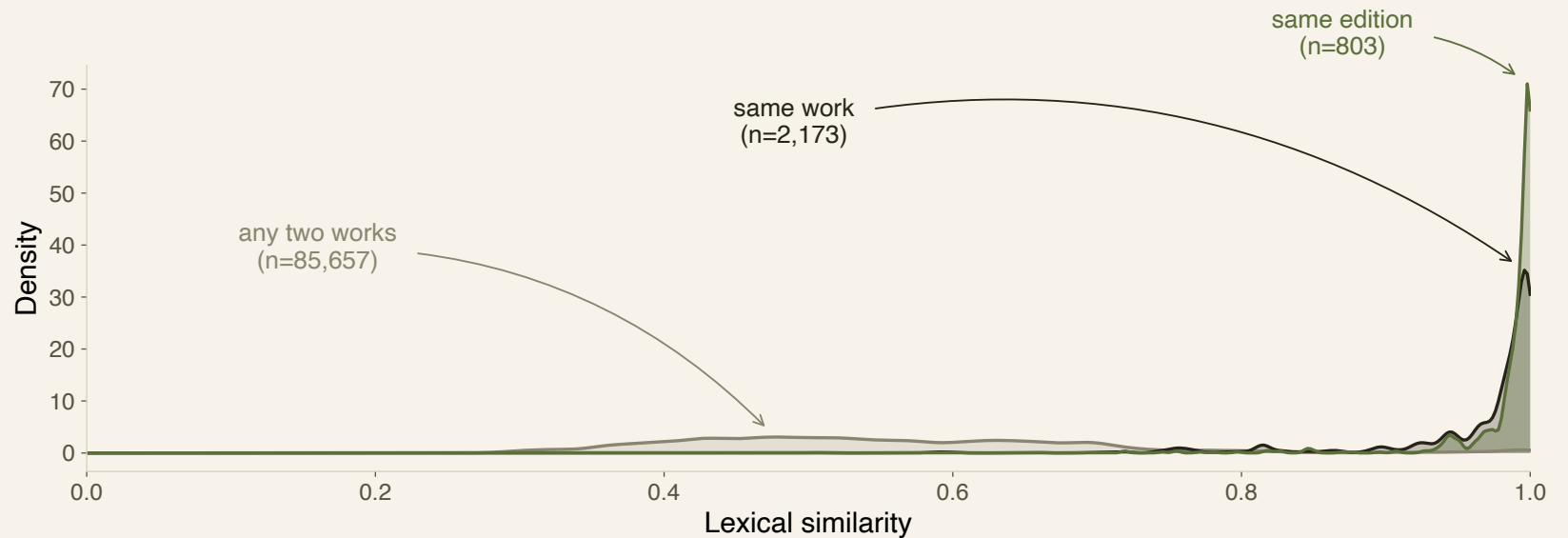
Curated fields are auditable

value + reasoning + checked/updated timestamp

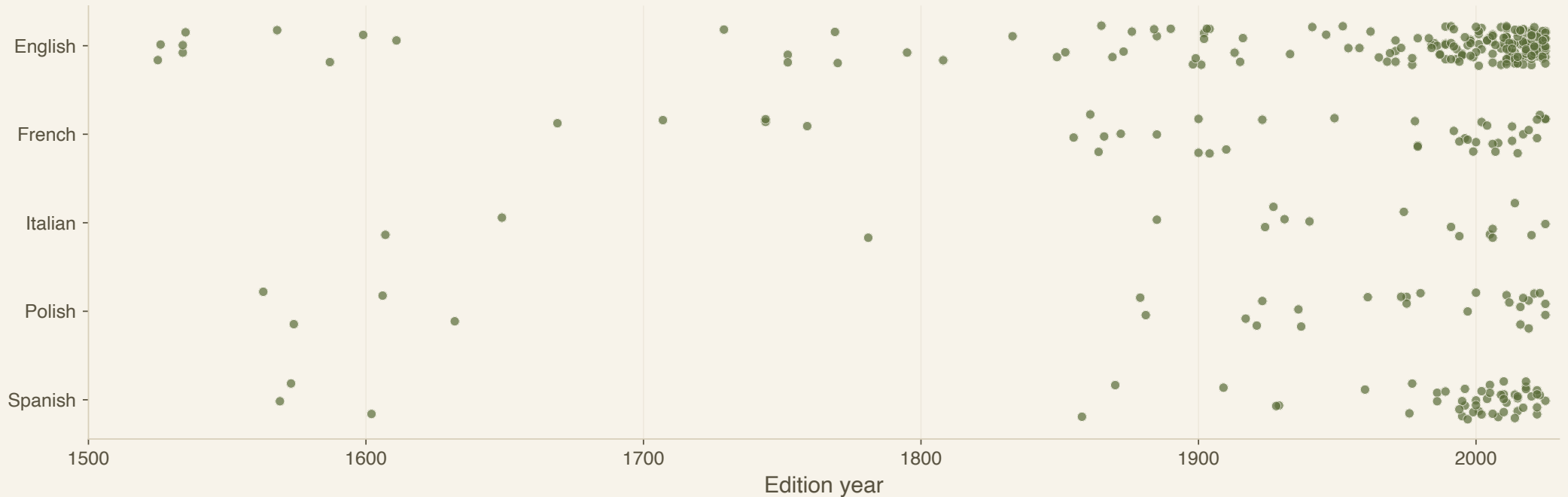
The corpus separates 651 source instances from 334 curated editions.

Three levels of translation similarity, pooled across five languages:

- **Same edition** from different sources → near-identical texts (~1.0): curation works
- **Same work**, different editions → wide spread; some editions are genuinely different texts
- **Any two works** → the full intra-language diversity



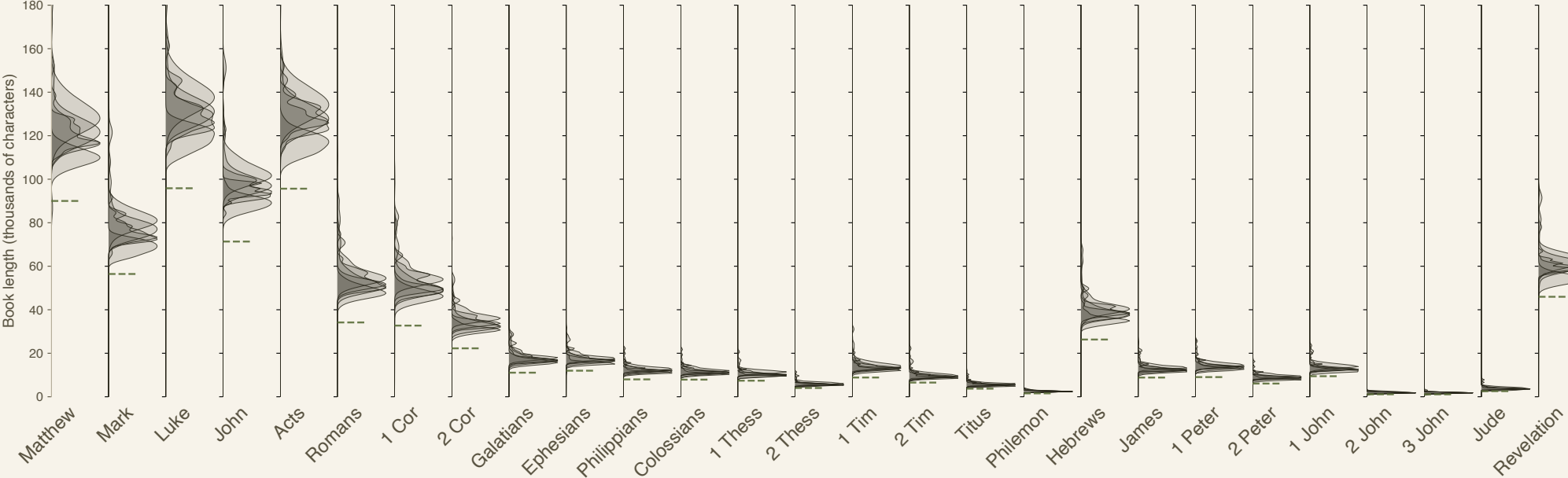
Diachronic Distribution of Translations



Each dot is a unique translation edition, positioned by publication year.

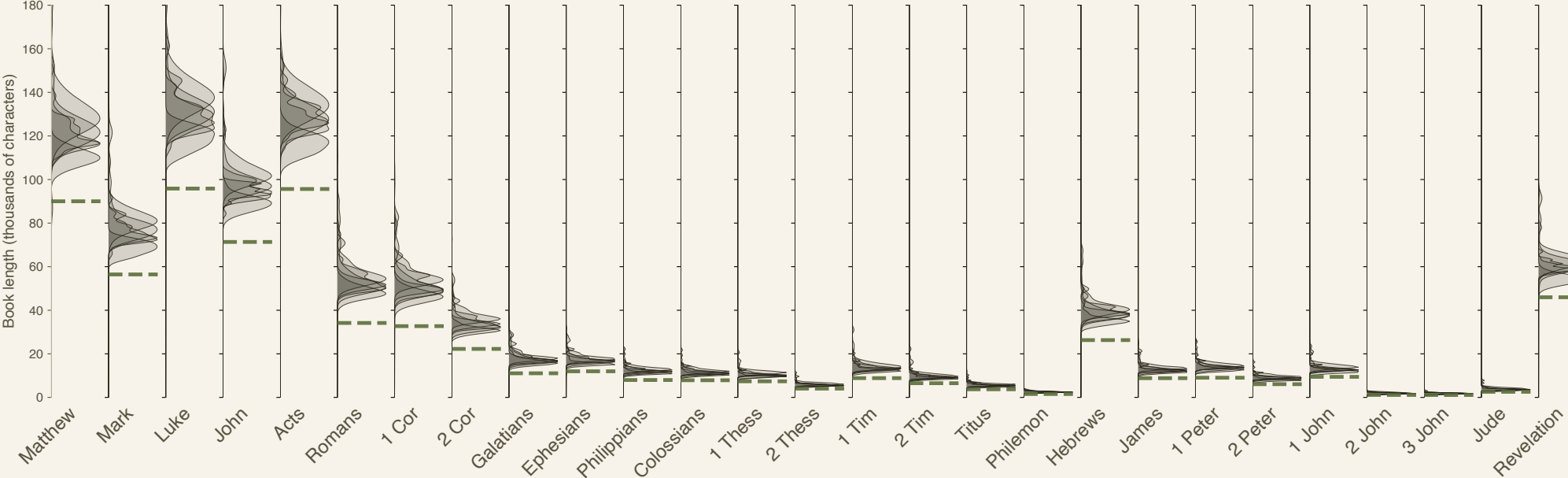
- The corpus spans texts from the **16th century** to the present.
- The densest period is the last few decades — translation history is still accelerating.

Book length distribution



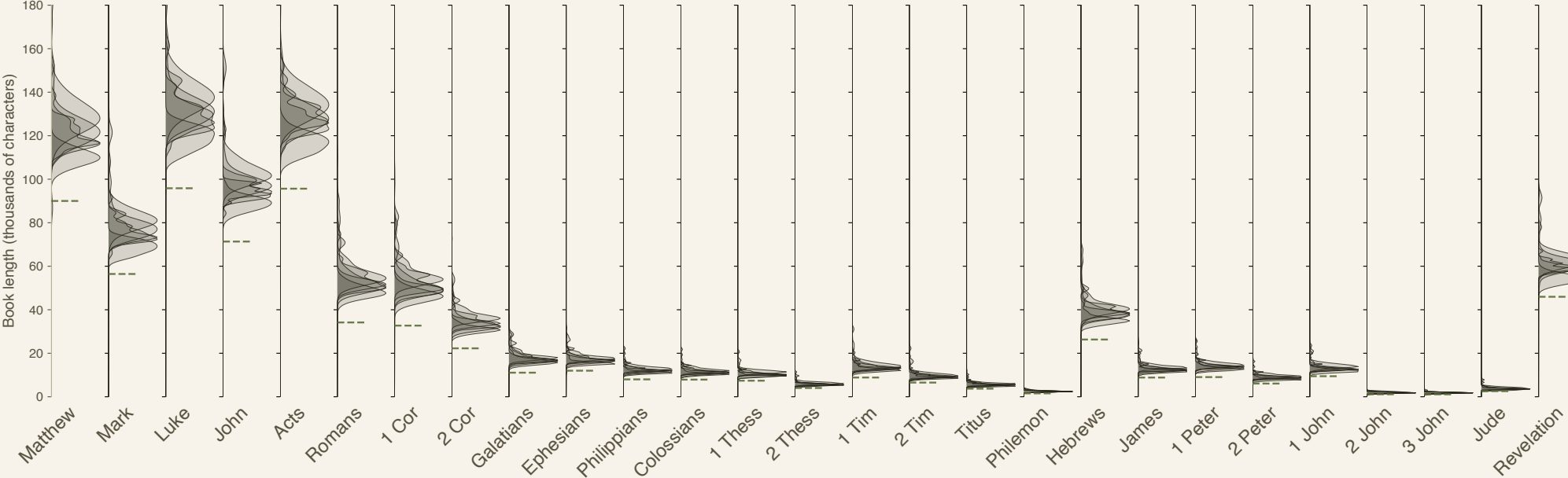
Distribution of book lengths across unique translations. Dashed line = Greek source (SBLGNT).

Book length distribution



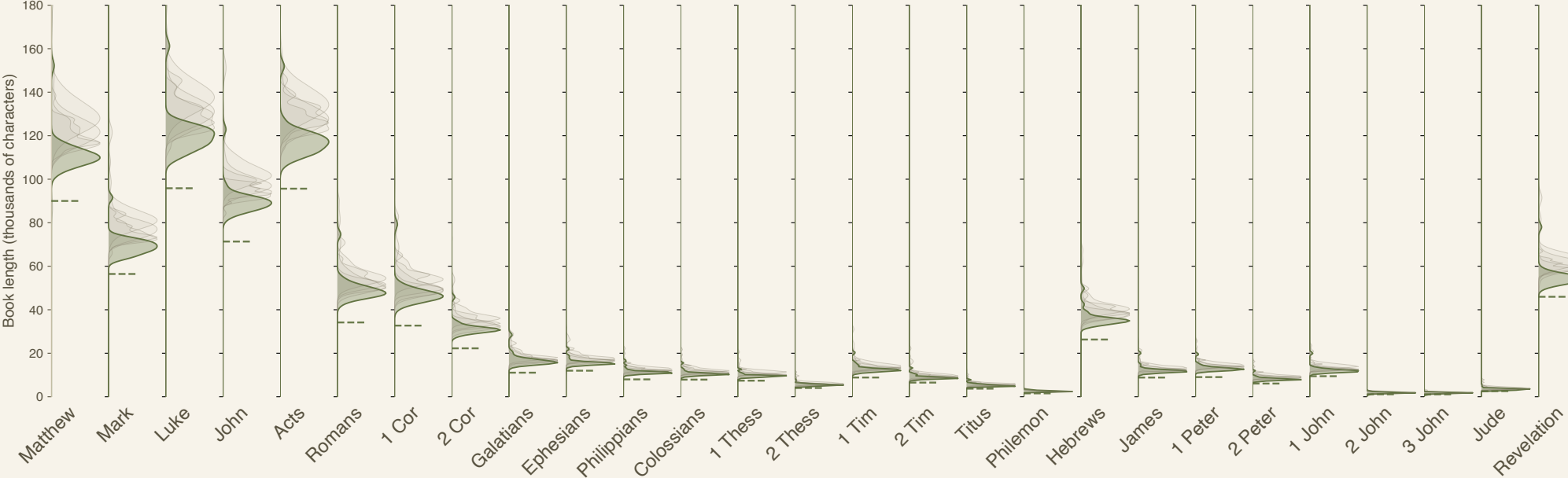
Distribution of book lengths across unique translations. Dashed line = Greek source (SBLGNT).

Book length distribution



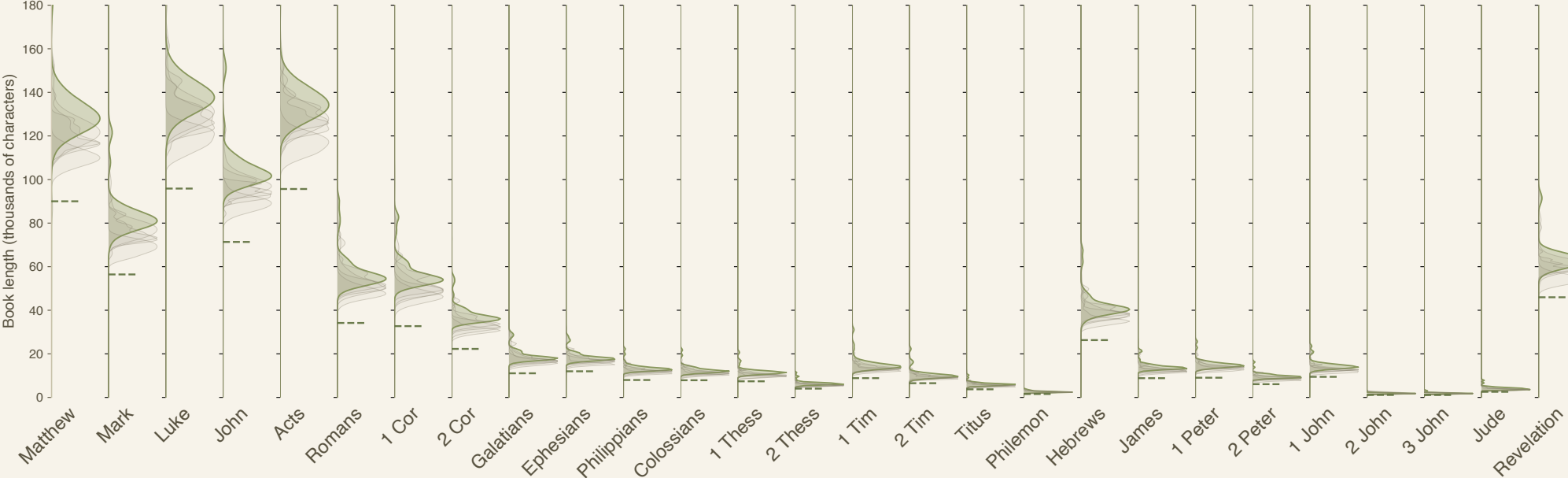
Distribution of book lengths across unique translations. Dashed line = Greek source (SBLGNT).

Book length distribution



Distribution of book lengths across unique translations. Dashed line = Greek source (SBLGNT).

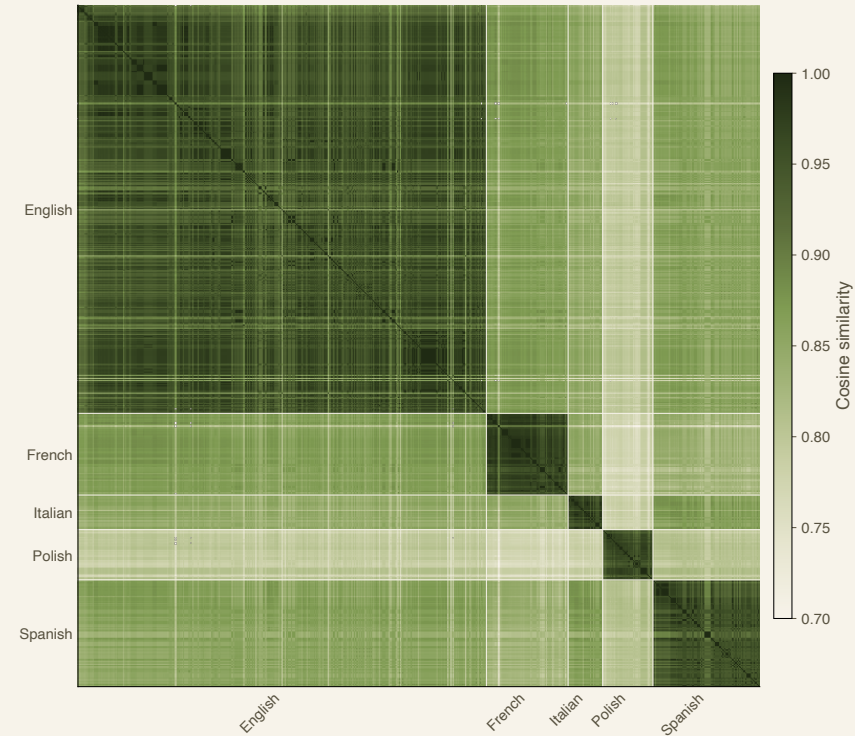
Book length distribution



Distribution of book lengths across unique translations. Dashed line = Greek source (SBLGNT).

Cross-Language Semantic Similarity

- Translations cluster by language first
- Romance languages form one block
- Polish is most distant; English sits in between



Pairwise semantic (Qwen3-Embedding-8B) similarity for all editions, chapter grain, sorted by language and year.

How can we measure the translator's intervention?

Prior work — the landscape

Prior work — the landscape

Evaluative

MT evaluation metrics (BLEU, METEOR, COMET, ...)

scalar

Quality estimation

scalar

Bias detection in MT/LLMs

predefined categories

Prior work — the landscape

Evaluative

MT evaluation metrics (BLEU, METEOR, COMET, ...)

scalar

Quality estimation

scalar

Bias detection in MT/LLMs

predefined categories

Categorical

Translation shifts / techniques

taxonomic

Translator style analysis

predefined features

Stylometry & authorship attribution

corpus-relative

Prior work — the landscape

Evaluative

MT evaluation metrics (BLEU, METEOR, COMET, ...)

scalar

Quality estimation

scalar

Bias detection in MT/LLMs

predefined categories

Categorical

Translation shifts / techniques

taxonomic

Translator style analysis

predefined features

Stylometry & authorship attribution

corpus-relative

Reductive

Translationese detection

binary

Literalness measurement

scalar

Style transfer & register conversion

generative

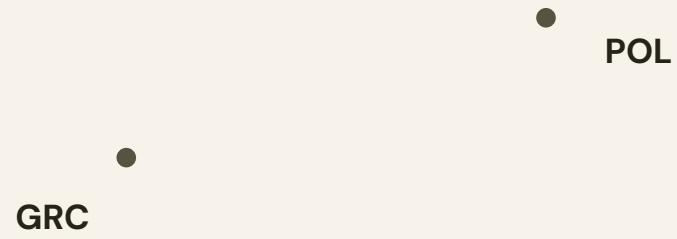
Prior work — the landscape

Evaluative	Categorical	Reductive
MT evaluation metrics (BLEU, METEOR, COMET, ...) scalar	Translation shifts / techniques taxonomic	Translationese detection binary
Quality estimation scalar	Translator style analysis predefined features	Literalness measurement scalar
Bias detection in MT/LLMs predefined categories	Stylometry & authorship attribution corpus-relative	Style transfer & register conversion generative

we need: continuous · directional · category-free · corpus-independent

Translation in vector space

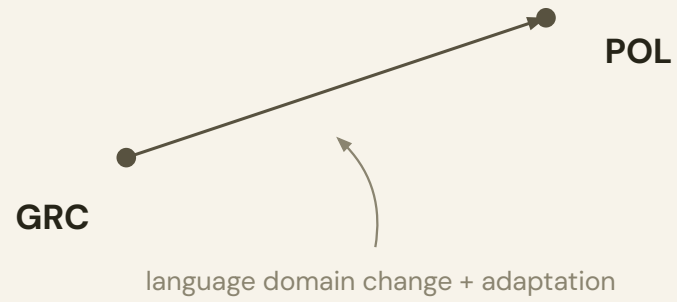
Translation in vector space



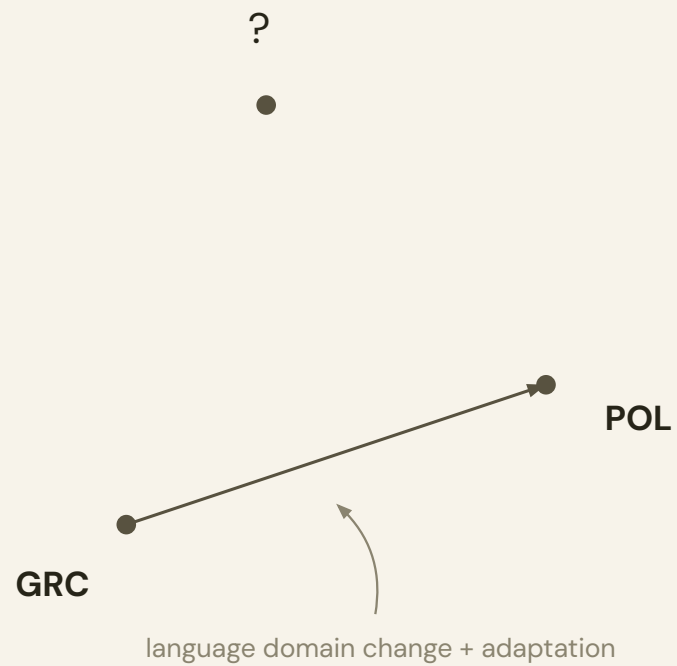
Translation in vector space



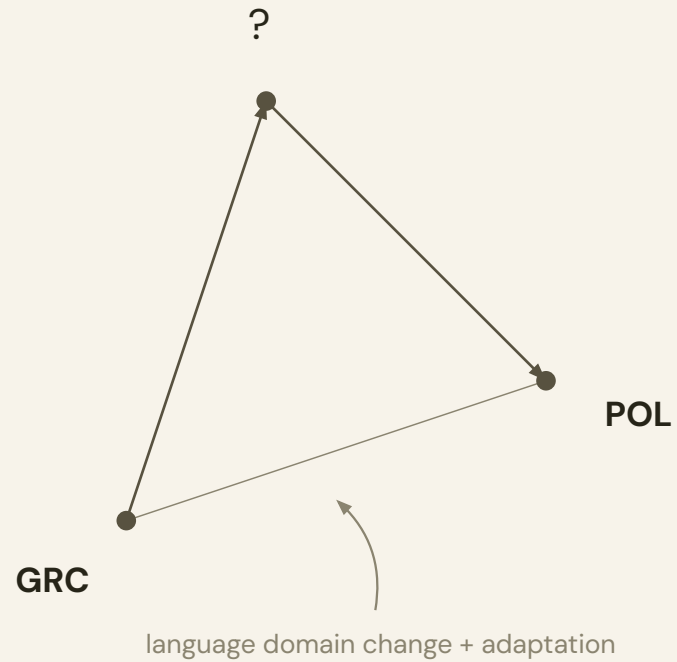
Translation in vector space



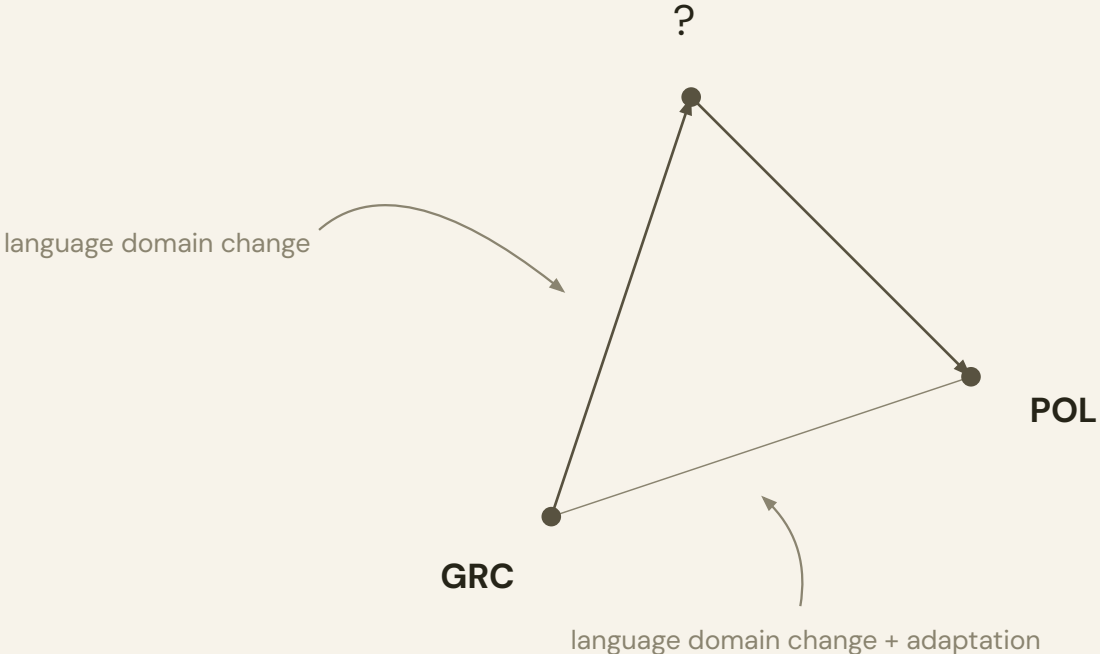
Translation in vector space



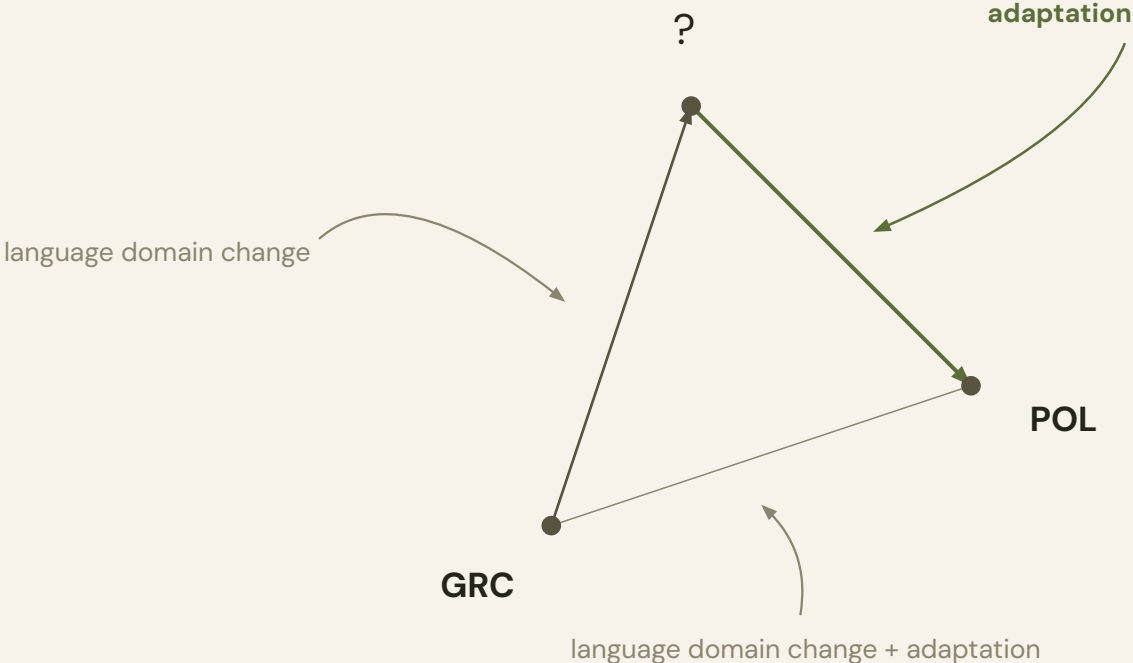
Translation in vector space



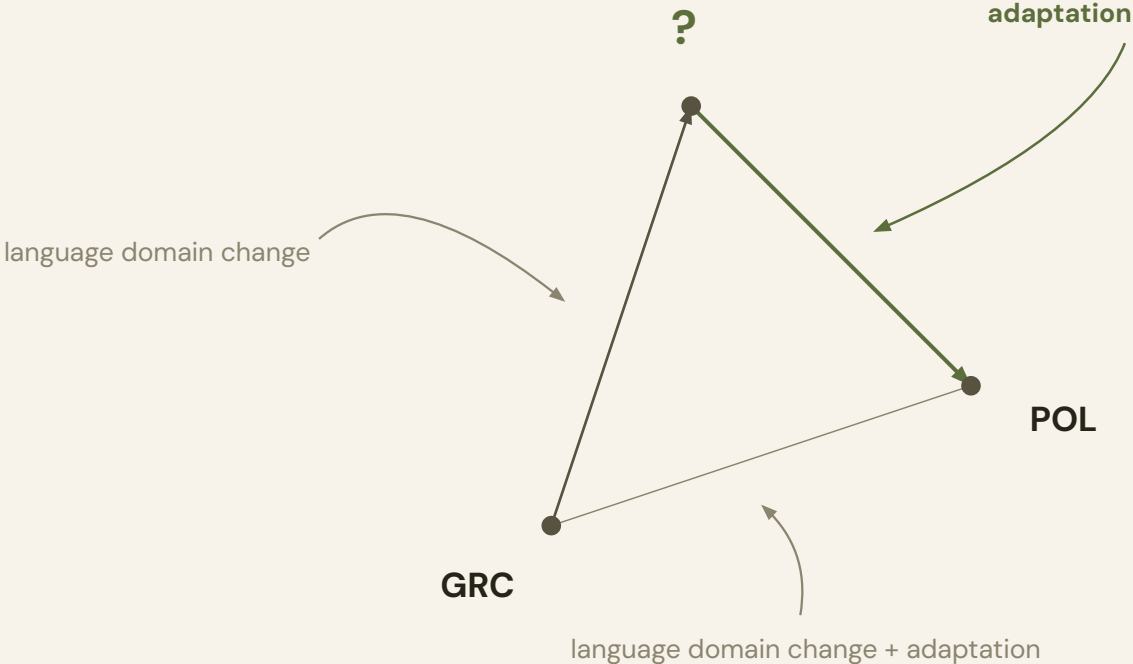
Translation in vector space



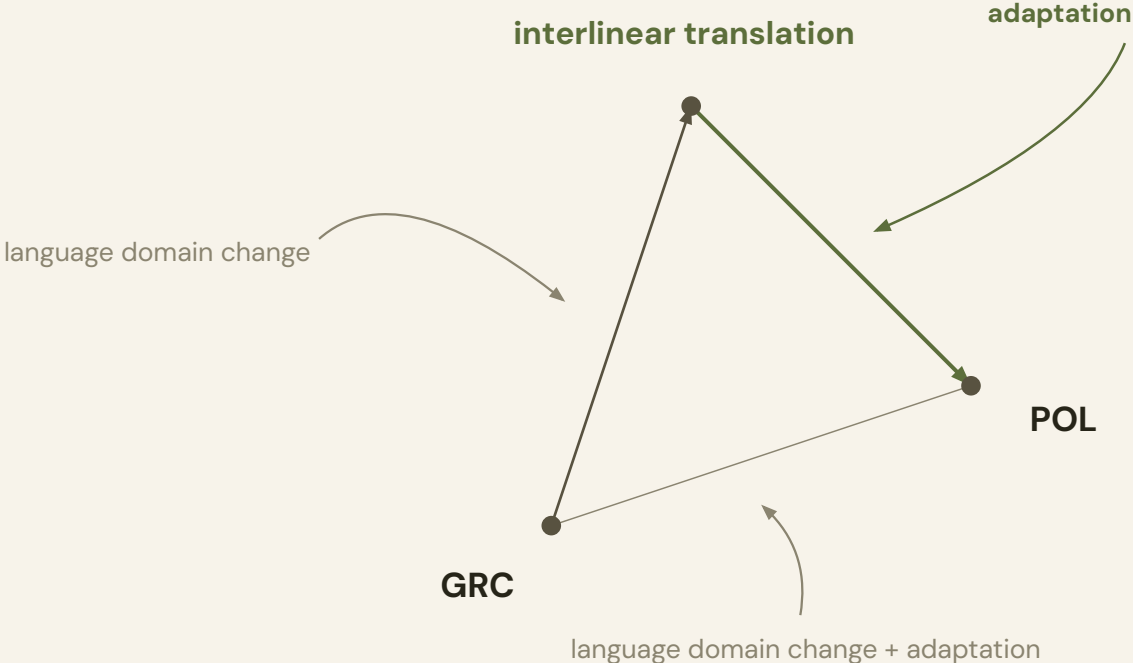
Translation in vector space



Translation in vector space



Translation in vector space



Interlinear Rendering Example

Ἐγένετο	ἄνθρωπος	ἀπεσταλμένος	παρὰ	θεοῦ	ὄνομα	αὐτῷ	Ἰωάννης
Became	a man	sent	from	God	name	to him	John
Stał się	człowiek	który jest wysłany	od	Boga	imię	mu	Jan

Fragment of an interlinear rendering into English and Polish (John 1:6).

Interlinear translation is a literal rendering of the source text in which each source-language word is aligned with a target-language counterpart, with as little syntactic restructuring as possible.

Ἐγένετο	ἄνθρωπος	ἀπεσταλμένος	παρὰ	θεοῦ	ὄνομα	αὐτῷ	Ἰωάννης
Became	a man	sent	from	God	name	to him	John
Stał się	człowiek	który jest wysłany	od	Boga	imię	mu	Jan

Fragment of an interlinear rendering into English and Polish (John 1:6).

The Intervention Vector

The intervention vector is the difference between the embedding of a literary translation and the embedding of its interlinear counterpart.

Define two difference vectors relative to the Greek source embedding:

$$\mathbf{v}_{\text{int}} = E(\text{INT}) - E(\text{GRC})$$

$$\mathbf{v}_{\text{lit}} = E(\text{LIT}) - E(\text{GRC})$$

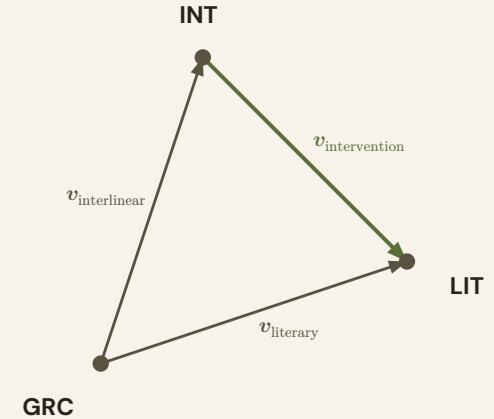
The intervention vector is their difference:

$$\mathbf{v}_{\text{interv}} = \mathbf{v}_{\text{lit}} - \mathbf{v}_{\text{int}}$$

After expansion, the Greek source cancels out:

$$\begin{aligned}\mathbf{v}_{\text{interv}} &= (E(\text{LIT}) - E(\text{GRC})) - (E(\text{INT}) - E(\text{GRC})) \\ &= E(\text{LIT}) - E(\text{INT})\end{aligned}$$

GRC Ἐγένετο ἄνθρωπος ἀπεσταλμένος...
INT *Became a man sent...*
LIT *There appeared a man sent...*



Source, interlinear, and literary embeddings for the same verse.

Experimental Setup

Validation of the intervention–vector method on a strategy–annotated subset of the Targum Corpus.

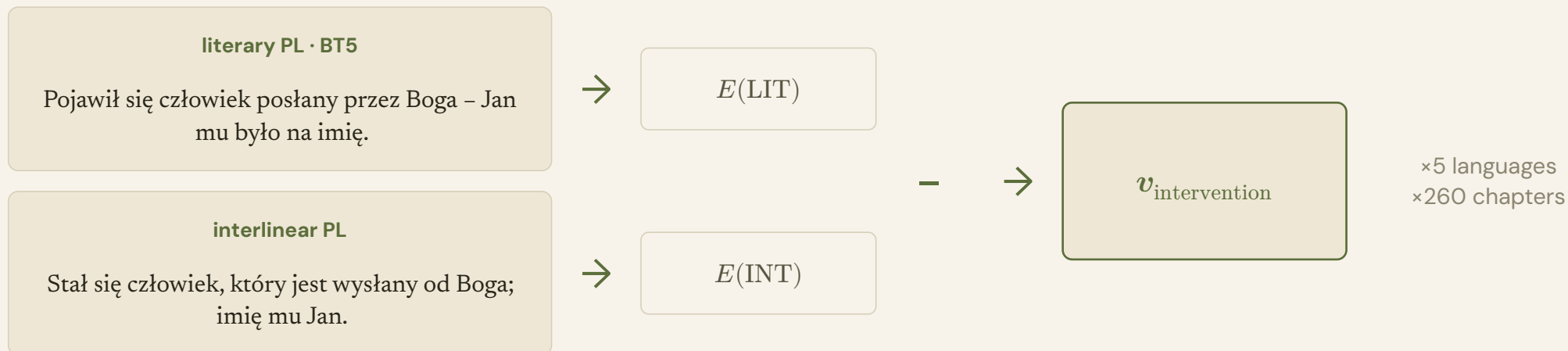
- **74 literary translations** in 5 languages (EN, FR, IT, PL, ES)
× **260 NT chapters**
- **5 interlinear translations** (one per language)
- **Embedding model:** Qwen3-Embedding-8B (chapter–level)

Each literary translation is manually classified on the scale **literal** → **formal** → **dynamic** → **paraphrase**.

Strategy	Eng.	Fr.	It.	Pol.	Sp.	Total
Literal	3	1	4	4	3	15
Formal	4	6	4	7	2	23
Dynamic	7	4	2	2	5	20
Paraphrase	2	3	2	3	6	16
Total	16	14	12	16	16	74

Strategy breakdown for 74 literary translations across 5 languages.

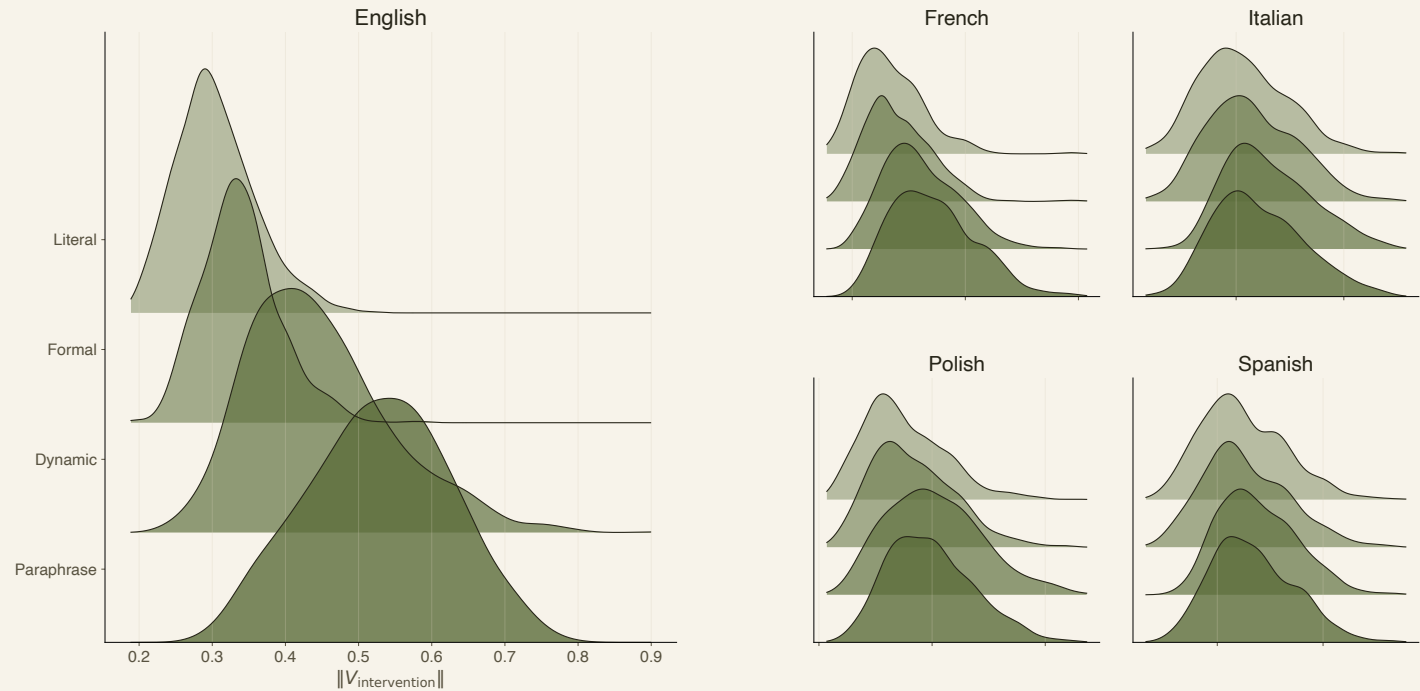
Intervention-Vector Pipeline



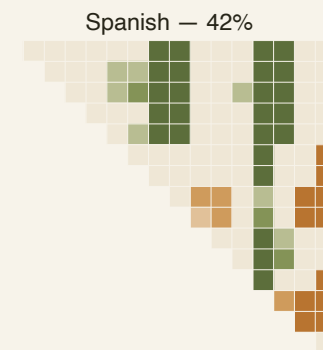
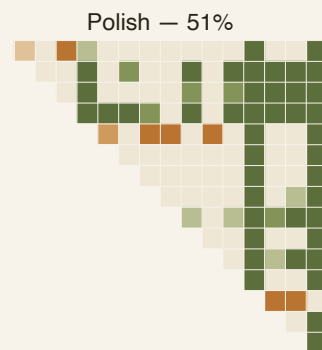
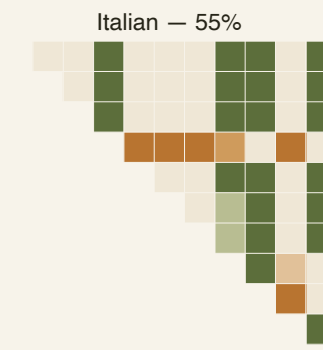
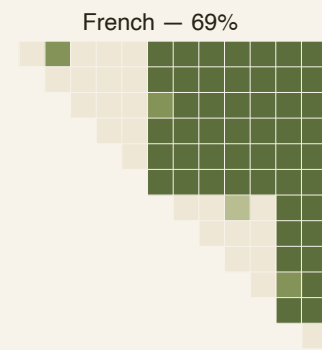
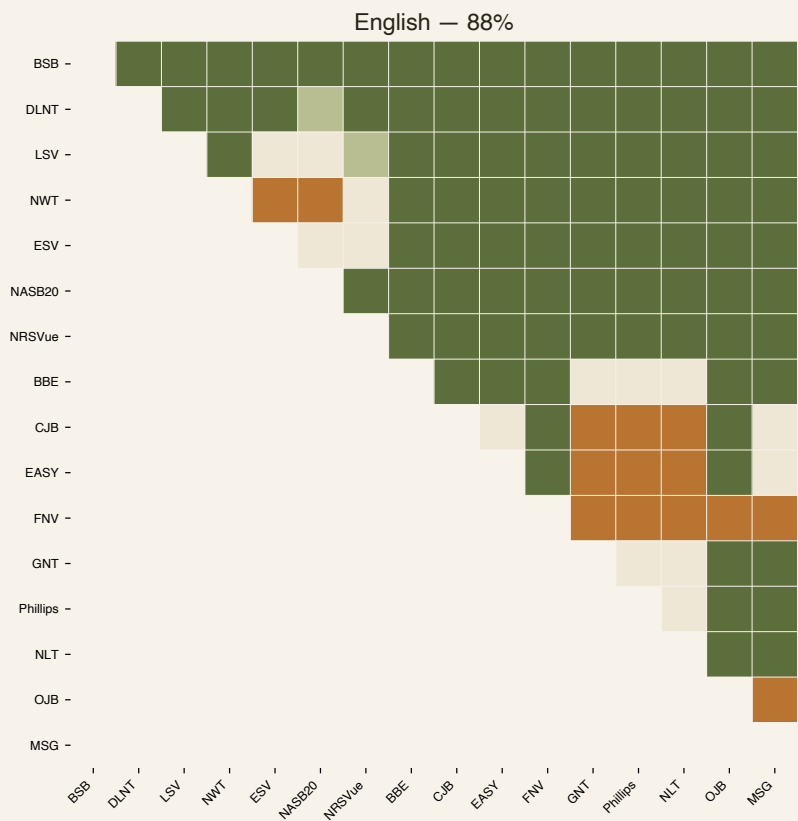
Intervention vector magnitude recovers the theoretical spectrum of translation strategy without supervision.

The more dynamic the strategy, the longer the intervention vector.

Magnitude does not yet tell us **how** the translator intervenes — it mainly tells us **how much** changes relative to the interlinear point.



Distributions of intervention vector magnitude by strategy, across 5 languages.



Pairwise Mann-Whitney tests over chapter-level intervention magnitudes. Upper triangle only; green cells mean row closer to INT, brown cells mean column closer. Percentages use unordered pair counts.

Magnitude Is Not Direction

LIT₁ · BT5

Pojawił się człowiek posłany przez Boga - Jan mu było na imię.

INT

Stał · się · człowiek · który · jest · wysłany · od · Boga · imię · mu · Jan

LIT₂ · WARSZ.PRASKA

Przyszedł człowiek posłany przez Boga. Jan miał na imię.

Magnitude Is Not Direction

LIT₁ · BT5
Pojawił się człowiek posłany przez Boga - Jan mu było na imię.



INT
Stał się człowiek który jest wysłany od Boga imię mu Jan

Qwen3-Embedding-8B



LIT₂ · WARSZ.PRASKA
Przyszedł człowiek posłany przez Boga. Jan miał na imię.

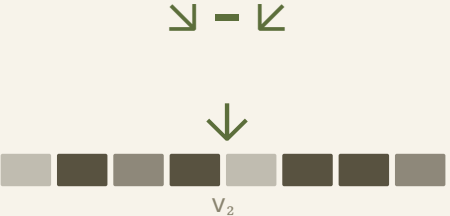
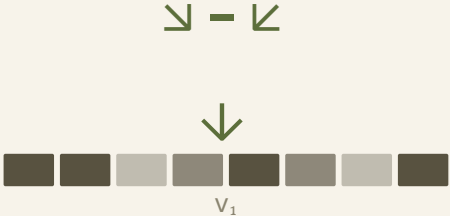
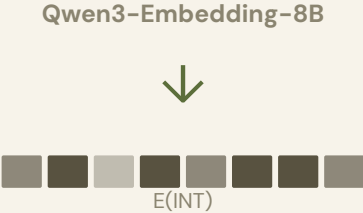


Magnitude Is Not Direction

LIT₁ · BT5
Pojawił się człowiek posłany przez Boga - Jan mu było na imię.

INT
Stał się człowiek · który · jest · wysłany · od · Boga · imię · mu · Jan

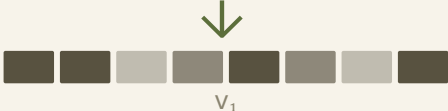
LIT₂ · WARSZ.PRASKA
Przyszedł człowiek posłany przez Boga. Jan miał na imię.



Magnitude Is Not Direction

LIT₁ · BT5

Pojawił się człowiek posłany przez Boga - Jan mu było na imię.



L2 norm

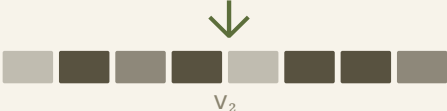


0.34

INT

Stał · się · człowiek · który · jest · wysłany · od · Boga · imię · mu · Jan

Qwen3-Embedding-8B



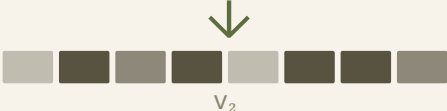
L2 norm



0.36

LIT₂ · WARSZ.PRASKA

Przyszedł człowiek posłany przez Boga. Jan miał na imię.

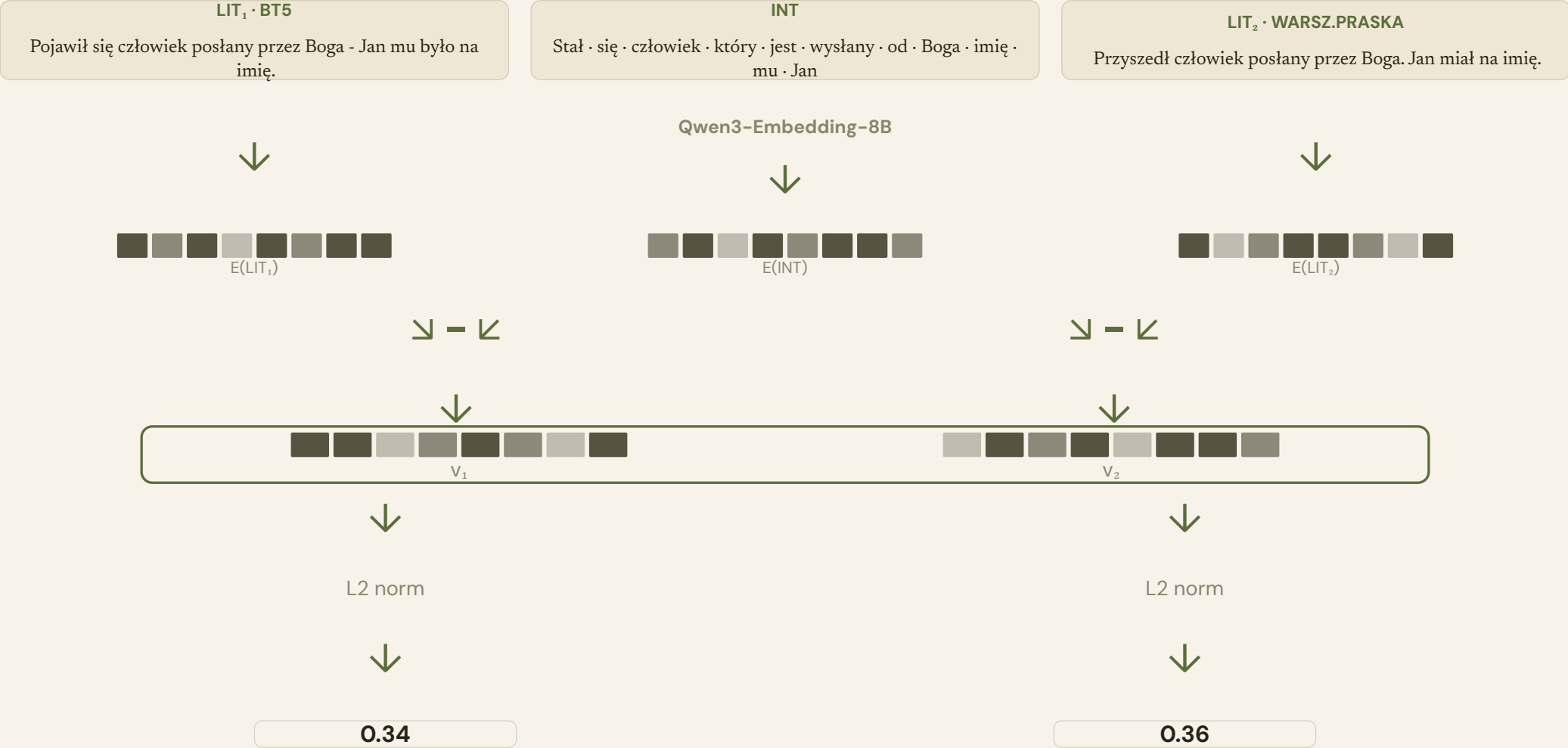


L2 norm



0.36

Magnitude Is Not Direction





PCA of chapter-level intervention vectors by language.

Same Magnitude, Different Direction

Both translations diverge from the interlinear translation significantly, but they move away in different directions: one domesticates, the other foreignizes.

Both translations have a large vector magnitude, but they move through different regions of the PCA space

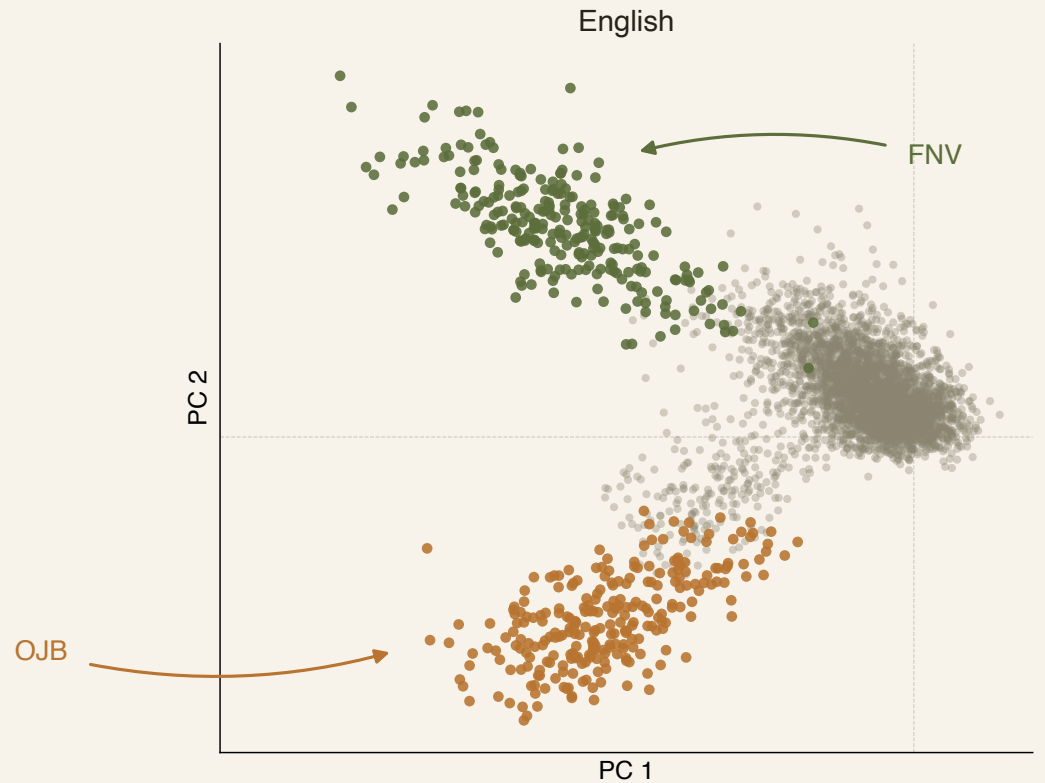
Acts 17:33

GRK Οὕτως ὁ Παῦλος ἐξῆλθεν ἐκ μέσου αὐτῶν.

INT Thus Paul went out from the midst of them.

FNV *So Small Man went on his way.*

OJB *Thus did Rav Sha'ul go out from the midst of them.*



English PCA facet with FNV/OJB clusters highlighted.

Targum: Richer Metadata

critical edition

Nestle–Aland 28

Textus Receptus

Byzantine text

source–text basis for interlinear and literary comparisons

Targum: Richer Metadata

critical edition

Nestle–Aland 28

Textus Receptus

Byzantine text

source–text basis for interlinear and literary comparisons

translation lineage

KJV → ASV → RSV

Reina–Valera revisions

Vulgate–mediated

families and revisions grouped as one intellectual translation work

Targum: Richer Metadata

critical edition

Nestle–Aland 28

Textus Receptus

Byzantine text

source–text basis for interlinear and literary comparisons

translation lineage

KJV -> ASV -> RSV

Reina–Valera revisions

Vulgate–mediated

families and revisions grouped as one intellectual translation work

confessional tradition

Catholic

Protestant

Orthodox

Ecumenical

institutional and doctrinal contexts recorded as analyzable metadata

Targum Interface

Targum תרגום Browse GitHub

A multilingual corpus of New Testament translations

1 CORINTHIANS 13:12
 Βλένομεν γὰρ ἀρτι δι' ὀμίχρου ἐν αἰνίγματι

Interlinear
 see-see · for · now · through · *mirror* · in · *riddle*

Young's Literal Translation
 * For we see now through a *mirror*, *obscurely* *

American Standard Version
 * For now we see in a *mirror*, *darkly* *

King James Version
 * For now we see through a *glass*, *darkly* *

Darby
 * For we see now through a *dim window* *obscurely* *

World English Bible
 * For now we see in a *mirror*, *dimly* *

651 instances of 334 editions across five languages, with verified canonical metadata.

● English ● French ● Italian ● Polish ● Spanish

Each dot is one edition, 334 editions, 5 languages, 1525-2025.

223 works, 334 editions, 651 instances.

Prior corpora typically pick one canonical text per translation family. We keep every edition we can verify – the 1611 King James and the 1769 Cambridge revision are different texts, and a corpus that collapses them loses the diachronic signal that makes within-family comparison interesting.

26% of works carry more than one edition, but those works account for 35% of instances.

Targum תרגום Browse GitHub

Browse the corpus

One row per canonical work. Type to filter – bare text matches names, or use fields like lang:pl year:2883.

Q *king james year:2003 · lang=en copyright:public · year:1500,1700*

223 works Sorted by year ↑

WORK	LANG	YEARS ↑	EDITIONS	COPYRIGHT	COVERAGE
Tyndale Bible tyndale-bible · View details -	EN	1525 – 2025	5	Public Domain	████████████████████
Coverdale Bible coverdale-bible · View details -	EN	1535	1	Public Domain	████████████████████
Biblia Brzeska biblia-brzeska · View details -	PL	1563	1	Public Domain	████████████████████
Bishops' Bible bishops-bible · View details -	EN	1568	1	Public Domain	████████████████████
Biblia del Oso biblia-del-oso · View details -	ES	1569	1	Public Domain	████████████████████
Biblia nieświeska szywon-bodny · View details -	PL	1574	1	Public Domain	████████████████████
Geneva Bible geneva-bible-eng · View details -	EN	1587 – 1599	2	Public Domain	████████████████████
Reina-Valera Antigua reina-valera-antigua · View details -	ES	1602	1	Public Domain	████████████████████
Nowy Testament rakowski rakowski-nt · View details -	PL	1606	1	Public Domain	████████████████████
Bibbia Diodati diodati-bible · View details -	IT	1607 – 1885	3	Public Domain	████████████████████
King James Version king-james-version · View details -	EN	1611 – 1962	5	Mixed	████████████████████
Biblia Gdanska biblia-gdanska · View details -	PL	1632 – 1881	2	Public Domain	████████████████████
Nouvelle Édition de Genève geneva-bible-fre · View details -	FR	1669 – 1979	2	Mixed	████████████████████
Bible Martin bible-martin · View details -	FR	1707 – 1744	2	Public Domain	████████████████████

Targum תרגום Browse GitHub

Biblia Brzeska

1563 · 1563
 Source: biblia@polska.pl · [Public domain](#)

READER

BOOK CHAPTER
 BROWSE **NAT** **1** GO TO

[Matthew 2](#) →

1 Toć iest wypisanie narodu Iezusa Krystusa syna Dawidowego, syna Abrahamowego. ² Z Abrahamo poszedł Izaak, z Izaaka poszedł Iakob, a z Iakoba poszedł Iuda y bracia jego. ³ Z Iudy poszedł Fares y Zara z Tamar, z Faresa poszedł Ezrom, z Ezroma poszedł Aram. ⁴ Z Arama poszedł Aminadab, z Aminadaba poszedł Naasson, z Naassona poszedł Salmon. ⁵ Z Salomona poszedł Booz z Rahab, z Boozo poszedł Obed z Rut, a z Obeda poszedł Iesse. ⁶ Z Iessego poszedł krol Dawid, a z Dawida Krola poszedł Salomon, z tey kora byla przed tym żoną Wryaszową. ⁷ Z Salomona poszedł Roboam, z Roboama poszedł Abia, z Abie poszedł Asa. ⁸ Z Asy poszedł Iozafat, a z Iozafata poszedł Ioram, a z Iorama poszedł Ozyasz. ⁹ Z Ozyasza poszedł Ioaam, z Ioaama poszedł Achas, z Achasa poszedł Ezechiasz. ¹⁰ Z Ezechiasza poszedł Manasses, z Manasses poszedł Amron, z Amona poszedł Iozyasz. ¹¹ Z Iozyasza poszedł Ioaqim, a z Ioaqima poszedł Iechoniasz y bracia iego, gdy byli zaprowadzeni do Babilonu. ¹² A po zaprowadzeniu do Babilonu, z Iechoniasza poszedł Salatyel, a z Salatyela poszedł Zorobabel. ¹³ Z Zorobabela poszedł Abiud, z Abiuda poszedł Eliakim, z Eliakima poszedł Azor. ¹⁴ Z Azora poszedł Sadok, z Sadoka poszedł Achin, z Achina poszedł Eliud. ¹⁵ Z Eliudu poszedł Eleazar, z Eleazar poszedł Matan, a z Matana poszedł Iakob. ¹⁶ Z Iakoba poszedł Iozef mąż Maryey, z korey się narodził Iezus, który iest narodem Krystus. ¹⁷ A tak uwieczniono pokolenie Iezusa, który iest narodem Krystus.

Interlinear Baselines: Next Steps

- ① Extending experimental scope: broader validation beyond the current annotated dataset.
- ② Packaging the method as an accessible blackbox tool for obtaining embeddings and intervention vectors.



mrpacz.com/ipipan2026

The research was supported by the National Science Centre, Poland under project number 2025/57/N/HS2/04961. We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Center: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2026/019145.