Interlinear Translation of Ancient Greek Texts

How Morphological Tags Enhance Machine Translation Quality

Maciej Rapacz | 2025-03-24

ἀπεκρίθη αὐτῷ ὁ ἀσθενῶν· Κύριε, ἄνθρωπον οὐκ ἔχω ἵνα ὅταν ταραχθῇ τὸ ὕδωρ βάλῃ με εἰς τὴν κολυμβήθραν· ἐν ᾧ δὲ ἔρχομαι ἐγὼ ἄλλος πρὸ ἐμοῦ καταβαίνει.

Odpowiedział Mu chory: «Panie, nie mam człowieka, aby mnie wprowadził do sadzawki, gdy nastąpi poruszenie wody. Gdy ja sam już dochodzę, inny wchodzi przede mną». (BT5)

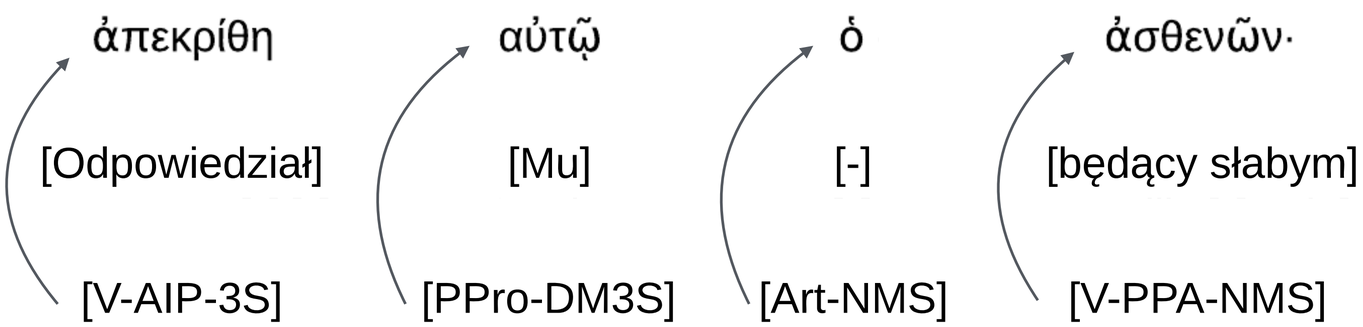
Odpowiedział mu on chory: Panie! nie ma człowieka, który by mię, gdy bywa poruszona woda, wrzucił do sadzawki; ale gdy ja idę, inszy przede mną wstępuje. (BG)

Chory odpowiedział: Panie, nie mam człowieka, który by mnie tuż po poruszeniu wody wrzucił do sadzawki, a zanim ja sam dojdę, inny wchodzi przede mną. (NP3)

Chory mu odpowiedział: Panie, nie mam człowieka, który wniósłby mnie do sadzawki, gdy woda zostaje poruszona. Lecz gdy ja idę, inny wchodzi przede mną. (UBG)

[Odpowiedział] [Mu] [-] [będący słabym] [Panie] [człowieka] [nie] [mam] [aby] [gdy] [zostałaby poruszona] [-] [woda] [wrzuciłby] [mnie] [do] [-] [sadzawki] [zanim] [gdy] [zaś] [przychodzę] [ja] [inny] [przede] [mną] [schodzi]

ἀπεκρίθη αὐτῷ ὁ ἀσθενῶν· Κύριε, ἄνθρωπον οὐκ ἔχω ἵνα ὅταν ταραχθῇ τὸ ὕδωρ βάλῃ με εἰς τὴν κολυμβήθραν· ἐν ᾧ δὲ ἔρχομαι ἐγὼ ἄλλος πρὸ ἐμοῦ καταβαίνει.



[Odpowiedział] [Mu] [-] [będący słabym] [Panie] [człowieka] [nie] [mam] [aby] [gdy] [zostałaby poruszona] [-] [woda] [wrzuciłby] [mnie] [do] [-] [sadzawki] [zanim] [gdy] [zaś] [przychodzę] [ja] [inny] [przede] [mną] [schodzi]

ἀπεκρίθη αὐτῷ ὁ ἀσθενῶν· Κύριε, ἄνθρωπον οὐκ ἔχω ἵνα ὅταν ταραχθῇ τὸ ὕδωρ βάλῃ με εἰς τὴν κολυμβήθραν· ἐν ᾧ δὲ ἔρχομαι ἐγὼ ἄλλος πρὸ ἐμοῦ καταβαίνει.

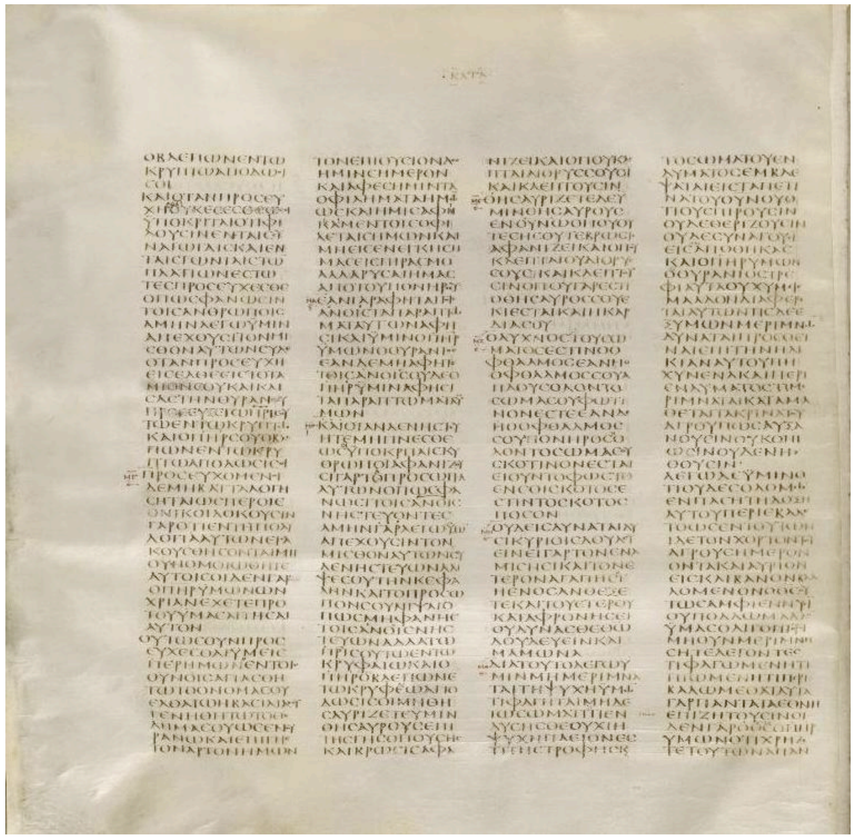
ἀπεκρίθη	αὐτῷ	ὁ	ἀσθενῶν·	
[Odpowiedział]	[Mu]	[-]	[będący słabym]	Chory odpowiedział:
[V-AIP-3S]	[PPro-DM3S]	[Art-NMS]	[V-PPA-NMS]	

[Odpowiedział] [Mu] [-] [będący słabym] [Panie] [człowieka] [nie] [mam] [aby] [gdy] [zostałaby poruszona] [-] [woda] [wrzuciłby] [mnie] [do] [-] [sadzawki] [zanim] [gdy] [zaś] [przychodzę] [ja] [inny] [przede] [mną] [schodzi]

Research Questions

- How well do the modern Machine Translation models perform on the interlinear translation task?
- Can morphological tags improve the quality of interlinear translation?
- How do specialized ancient language models compare to general multilingual models?
- What impact do text preprocessing methods have on translation performance?

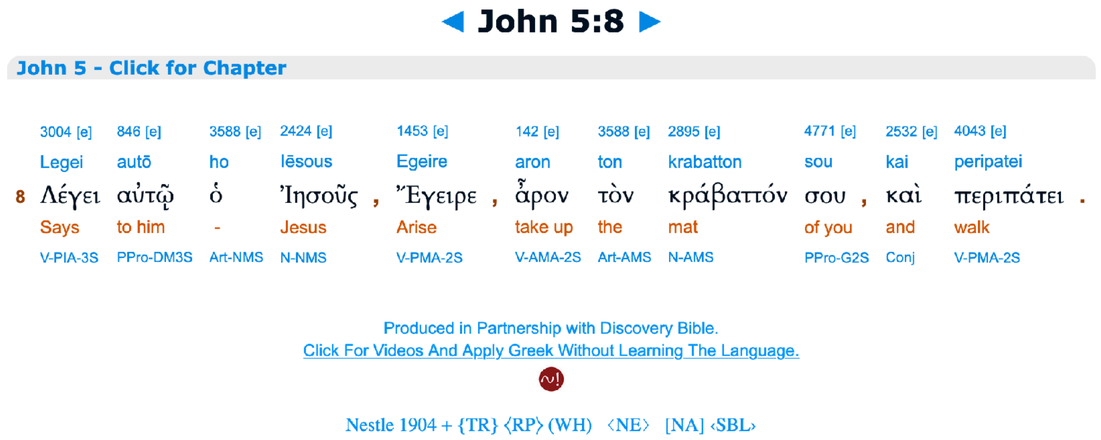
Dataset: Source Corpora



https://pl.wikipedia.org/wiki/Kodeks_Synajski

Dataset: Source Corpora

biblehub.com



◀ John 5:8 ▶

John 5 - Click for Chapter

	3004 [e]	846 [e]	3588 [e]	2424 [e]	1453 [e]	142 [e]	3588 [e]	2895 [e]	4771 [e]	2532 [e]	4043 [e]
	Legei	autō	ho	Iēsous	Egeire	aron	ton	krabatton	sou	kai	peripatei
8	Λέγει	αὐτῷ	ὁ	Ἰησοῦς ,	Ἔγειρε ,	ἆρον	τὸν	κράβαττόν	σου ,	καὶ	περιπάτει .
	Says	to him	-	Jesus	Arise	take up	the	mat	of you	and	walk
	V-PIA-3S	PPro-DM3S	Art-NMS	N-NMS	V-PMA-2S	V-AMA-2S	Art-AMS	N-AMS	PPro-G2S	Conj	V-PMA-2S

Produced in Partnership with Discovery Bible.
Click For Videos And Apply Greek Without Learning The Language.

Nestle 1904 + {TR} ⟨RP⟩ (WH) ‹NE› [NA] ‹SBL›

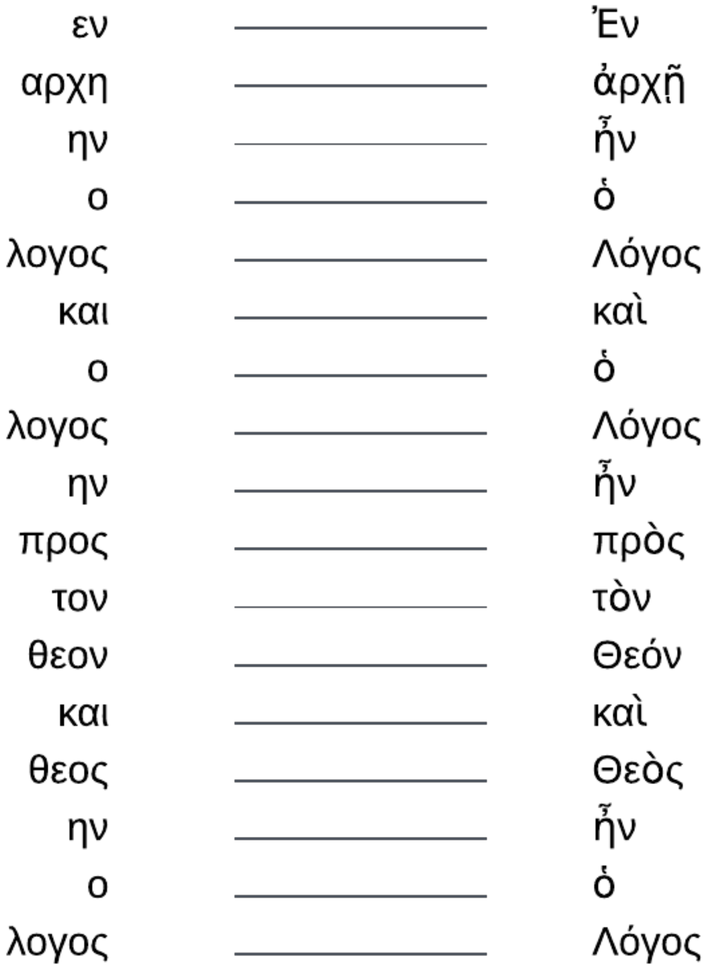
oblubienica.eu

nr	1	2	3	4	5	6	7	8	9	10	11
	3004	846	3588	2424	1453	142	3588	2895	4675	2532	4043
strong	λεγει	αυτω	ο	ιησους	εγειρε	αρον	τον	κραβαττον	σου	και	περιπατει
słowo	legei	autō	ho	iēsous	egeire	aron	ton	krabatton	sou	kai	peripatei
trans. fonet.	vi Pres Act 3 Sg	pp Dat Sg m	t_ Nom Sg m	n_ Nom Sg m	vm Pres Act 2 Sg	vm Aor Act 2 Sg	t_ Acc Sg m	n_ Acc Sg m	pp 2 Gen Sg	Conj	vm Pres Act 2 Sg
kod gramat.	Mówi	mu		Jezus	wstawaj	weź		matę	swoją	i	chodź
znaczenie				[JHWH jest zbawieniem]	podnoś się	zabierz		posłanie			
drugie znaczenie											

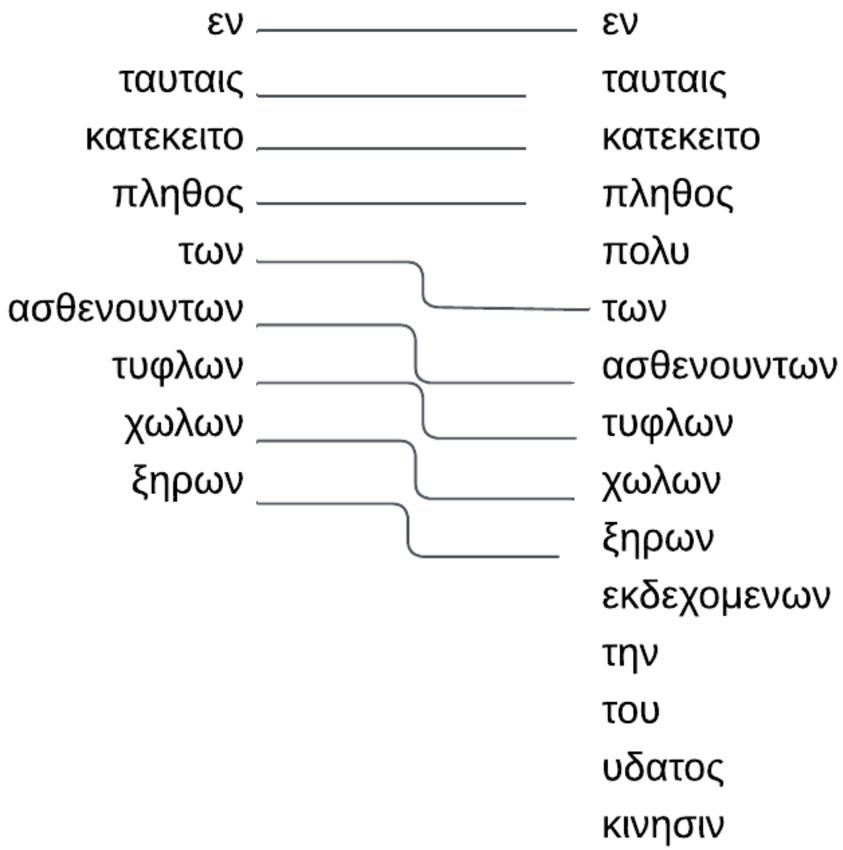
Wstecz Dalej

Portal oblubienica.eu

Dataset: Alignment



Dataset: Alignment



Bible Hub

Greek Text:

Base Text:

Nestle 1904 ⟨NE⟩
Eberhard Nestle, *H ΚΑΙΝΗ ΔΙΑΘΗΚΗ. Text with Critical Apparatus*. (British and Foreign Bible Society, 1904).
https://sites.google.com/ site/nestle1904

Variant Texts:

SBLGNT ‹SBL›
Michael W. Holmes, *Greek New Testament: SBL Edition*. (Society of Biblical Literature, 2010).

NA27 [NA]
Nestle-Aland Novum Testamentum Graece. 27th ed. Stuttgart: (Deutsche Bibelgesellschaft, 1993).

Westcott and Hort (WH)
Brooke Foss Westcott and Fenton John Anthony Hort, *The New Testament in the Original Greek*, vol. 1: *Text*; vol. 2: *Introduction* [and] *Appendix* (Cambridge: Macmillan, 1881).

Byzantine Majority Text ‹RP›
Maurice A. Robinson and William G. Pierpont, *The New Testament in the Original Greek: Byzantine Textform, 2005*. (Chilton Book Publishing, 2005).

Schrivener's Textus Receptus 1896 {TR}
F. H. A. Scrivener , *The New Testament in the Original Greek according to the Text followed in the Authorised Version* (Cambridge: University Press, 1894).

Oblubienica.eu

Praca opiera się na

- Textus Receptus (Tekście Przyjętym) - Scriveners Textus Receptus 1894 (Prepared and edited by Dr. Maurice A. Robinson.)
- Majority Text/Byzantium Text (Tekst Większościowy) - Byzantine Textform - Robinson & Pierpon 2005
- Nestle Aland Novum Testamentum Graece 28

	3778 [e]	1125 [e]	4771 [e]	2443 [e]	1492 [e]	3754 [e]	2222 [e]	2192 [e]	166 [e]		3588 [e]	4100 [e]	1519 [e]
	Tauta	egrapsa	hymin	hina	eidēte	hoti	zōēn	echete	aiōnion	,	tois	pisteuousin	eis
13	Ταῦτα	ἔγραψα	ὑμῖν	ἵνα	εἰδῆτε	ὅτι	ζωὴν	ἔχετε	αἰώνιον	,	τοῖς	πιστεύουσιν	εἰς
	These things	have I written	to you	so that	you may know	that	life	you have	eternal		to those	believing	in
	DPro-ANP	V-AIA-1S	PPro-D2P	Conj	V-RSA-2P	Conj	N-AFS	V-PIA-2P	Adj-AFS		Art-DMP	V-PPA-DMP	Prep

3588 [e]	3686 [e]	3588 [e]	5207 [e]	3588 [e]	2316 [e]		2532 [e]	2443 [e]	4100 [e]	1519 [e]	3588 [e]	3686 [e]	3588 [e]	5207 [e]	3588 [e]
to	onoma	tou	Huiou	tou	Theou	.	kai	hina	pisteuēte	eis	to	onoma	tou	huiou	tou
τὸ	ὄνομα	τοῦ	Υἱοῦ	τοῦ	Θεοῦ	.	⟨καὶ	ἵνα	~~πιστεύητε~~	~~εἰς~~	~~τὸ~~	~~ὄνομα~~	~~τοῦ~~	~~υἱοῦ~~	~~τοῦ~~
the	name	of the	Son	-	of God		and	that	you may believe	on	the	name	of the	Son	-
Art-ANS	N-ANS	Art-GMS	N-GMS	Art-GMS	N-GMS		Conj	Conj	V-PSA-2P	Prep	Art-ANS	N-NNS	Art-GMS	N-GMS	Art-GMS

2316 [e]	
theou	
~~θεοῦ~~⟩	.
of God	
N-GMS	

Wstecz Dalej Jana I 5 13 Przejdź do widoku rozdziałów

Nestle/Aland 28

- Textus/Receptus
- Majority Text/Byzantium Text (Tekst Większościowy)
- ✓ Nestle/Aland 28

tarze Przekład dosłowny Biblia Gdańska

Biblia Tysiąclecia King James Version (Ang)

nr	1	2	3	4	5	6
strong	5023	1125	5213	2443	1492	3754
słowo	ταυτα	εγραψα	υμιν	ινα	ειδητε	οτι
trans. fonet.	tauta	egrapsa	hymin	hina	eidēte	hoti
kod gramat.	pd Acc Pl n	vi Aor Act 1 Sg	pp 2 Dat Pl	Conj	vs Perf Act 2 Pl	Conj
znaczenie	Te	napisałem	wam	aby	wiedzielibyście	że
drugie znaczenie						

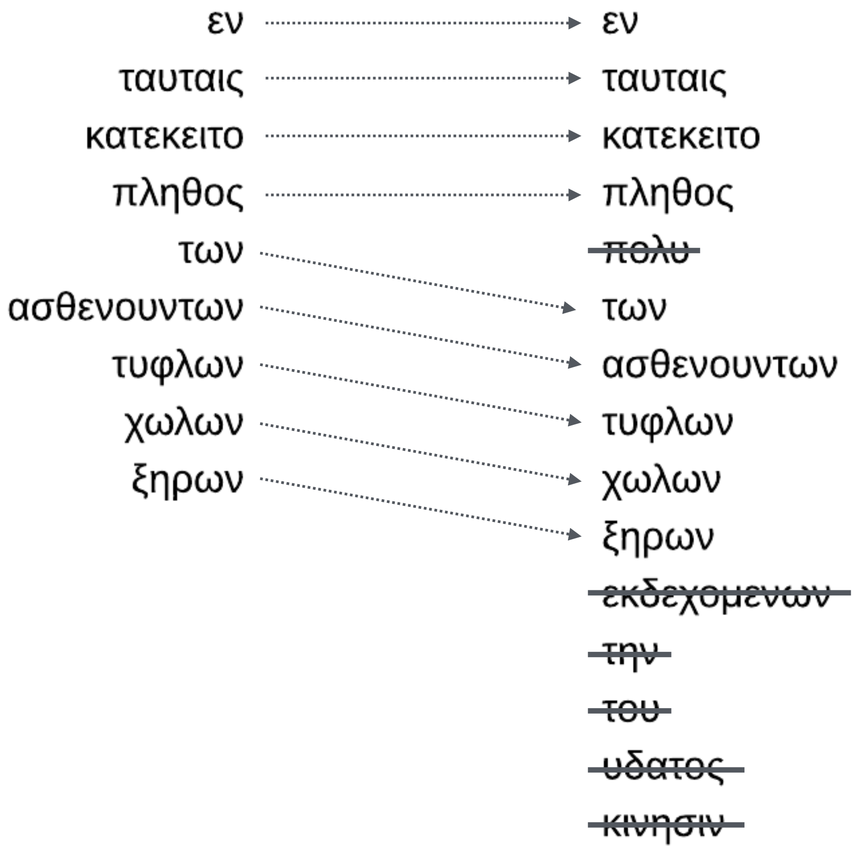
7	8	9	10	11	12	13
2222	2192	166	3588	4100	1519	3588
ζωην	εχετε	αιωνιον	τοις	πιστευουσιν	εις	το
zōēn	echete	aiōnion	tois	pisteuousin	eis	to
n_ Acc Sg f	vi Pres Act 2 Pl	a_ Acc Sg n	t_ Dat Pl m	vp Pres Act Dat Pl m	Prep	t_ Nom Sg n
życie	macie	wieczne		którzy wierzycie	w	

14	15	16	17	18
3686	3588	5207	3588	2316
ονομα	του	υιου	του	θεου
onoma	tou	hyiou	tou	theou
n_ Acc Sg n	t_ Gen Sg m	n_ Gen Sg m	t_ Gen Sg m	n_ Gen Sg m
imię		Syna		Boga

Wstecz Dalej

Portal oblubienica.eu

Dataset: Alignment



Dataset: Alignment

Prep	εν	.....	Ἐν	Prep
N-DFS	αρχη	.....	ἀρχῇ	n_ Dat Sg f
V-IIA-3S	ην	.....	ἦν	vi Impf vxx 3 Sg
Art-NMS	ο	.....	ὁ	t_ Nom Sg m
N-NMS	λογος	.....	Λόγος	n_ Nom Sg m
Conj	και	.....	καὶ	Conj
Art-NMS	ο	.....	ὁ	t_ Nom Sg m
N-NMS	λογος	.....	Λόγος	n_ Nom Sg m
V-IIA-3S	ην	.....	ἦν	vi Impf vxx 3 Sg
Prep	προς	.....	πρὸς	rep
Art-AMS	τον	.....	τὸν	_ Acc Sg m
N-AMS	θεον	.....	Θεόν	_ Acc Sg m
Conj	και	.....	καὶ	Conj
N-NMS	θεος	.....	Θεὸς	_ Nom Sg m
V-IIA-3S	ην	.....	ἦν	i Impf vxx 3 Sg
Art-NMS	ο	.....	ὁ	t_ Nom Sg m
N-NMS	λογος	.....	Λόγος	_ Nom Sg m

Two Tag Sets: Discrepancies

Two Tag Sets

- Total Unique Tags:
  - BibleHub: 684
  - Oblubienica: 1070

Grammatical Categories in the Corpora
Part of Speech: Verb, Noun, Adverb, Adjective, Article, Pronoun, Preposition, Conjunction, Interjection, Particle, Aramaic Word, Hebrew Word
Pronoun Subtype: Personal / Possessive, Demonstrative, Interrogative / Indefinite, Reciprocal, Relative and Reflexive
Person: 1st, 2nd, 3rd
Tense: Present, Imperfect, Future, Aorist, Perfect, Pluperfect
Mood: Indicative, Imperative, Subjunctive, Optative, Infinitive, Participle
Voice: Active, Middle, Passive, Middle or Passive
Case: Nominative, Vocative, Accusative, Genitive, Dative
Number: Singular, Plural
Gender: Masculine, Feminine, Neuter
Degree: Positive, Comparative, Superlative

Part of Speech	Bible Hub	Oblubienica
Verb	385	743
Pronoun	169	193
Adjective	68	56
Noun	31	39
Article	30	23
Adverb	3	5
Particle	3	4
Interjection	1	1
Preposition	1	1
Conjunction	1	1
Hebrew Word	1	1
Aramaic Word	0	1

Two Tag Sets

Discrepancies

- 2 Aorists Distinction
- **+100** unique forms

4905 [e]

synelthein

συνελθεῖν

coming together

V-ANA

18

4905

συνελθειν

synelthein

vn 2Aor Act

zejść się

1165 [e]

deigmatisai

δειγματίσαι ,

to expose publicly

V-ANA

12

1165

δειγματισαι

deigmatisai

vn Aor Act

wystawić na pokaz

wystawić jej na widok publiczny

Two Tag Sets

Discrepancies

- Marking Attic Dialect
- **+50** unique forms

1410 [e]

ēdynēthēsan

ἠδυνήθησαν

they were able

V-AIP-3P

9

1410

ηδυνηθησαν

ēdynēthēsan

vi Aor pasD 3 Pl Att

mogli

vi Aor pasD 3 Pl Att

Czasownik Tryb orzekający Czas przeszły dokonany Deponens bierny Trzecia osoba Liczba mnoga Attycka forma grecka

5463 [e]

echarēsan

ἐχάρησαν

they rejoiced [with]

V-AIP-3P

5

5463

εχαρησαν

echarēsan

vi 2Aor pasD 3 Pl

uradowali się

ucieszyli się

Two Tag Sets

Discrepancies

- Marking Deponent Forms
- **+200** unique forms

1096 [e]

egenomēn

ἐγενόμην

I became

V-AIM-1S

13

1096

εγενομην

egenomēn

vi 2Aor midD 1 Sg

stałem się

4388 [e]

proethemēn

προεθέμην

I purposed

V-AIM-1S

9

4388

προεθεμην

proethemēn

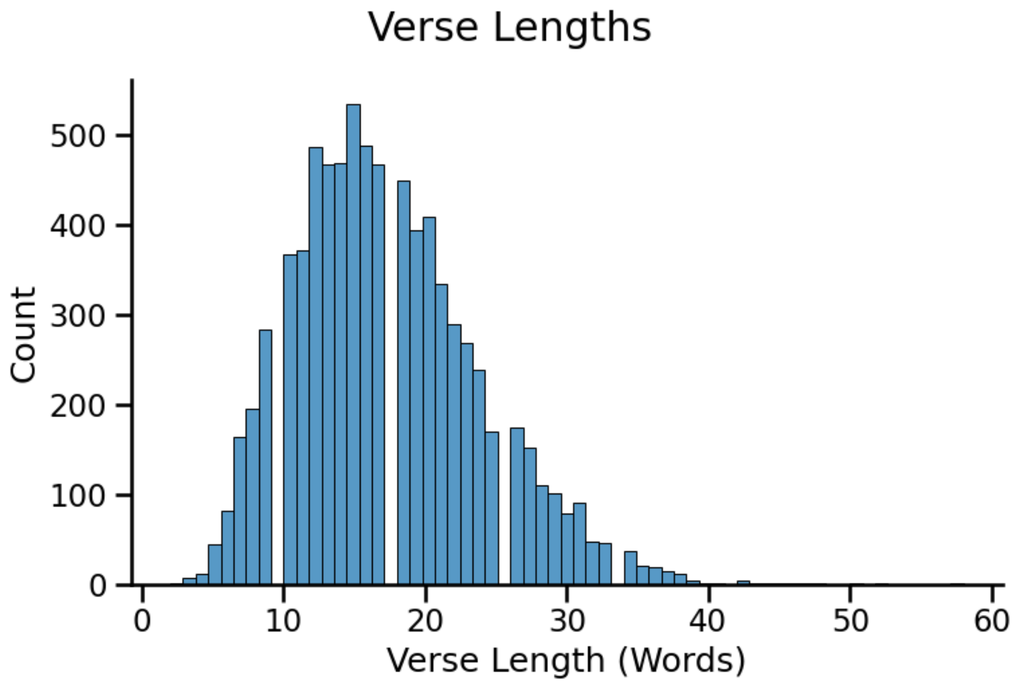
vi 2Aor Mid 1 Sg

postanowiłem

powziąłem sobie

Dataset: Final Stats

- Total
  - 7940 verses
  - 137k words
- 3 Splits (by verses)
  - Train: 80% (6352),
  - Val: 10% (794),
  - Test: 10% (794)
- 2 Tag Sets
  - BibleHub: 684 unique tags
  - Oblubienica: 1070 unique tags



Experiments

Methodology: Base Models

Exploring Large Language Models for Classical Philology

Frederick Riemenschneider
Dept. of Computational Linguistics
Heidelberg University
69120 Heidelberg
riemenschneider@cl.uni-heidelberg.de

Anette Frank
Dept. of Computational Linguistics
Heidelberg University
69120 Heidelberg
frank@cl.uni-heidelberg.de

mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer

Linting Xue* **Noah Constant*** **Adam Roberts***
Mihir Kale **Rami Al-Rfou** **Aditya Siddhant** **Aditya Barua** **Colin Raffel**
Google Research

bowphs/**GreTa**

Pre-trained on: Ancient Greek

bowphs/**PhilTa**

Pre-trained on:
Ancient Greek, Latin, English

T5 Architecture

google/**mt5**

Pre-trained on:
101 languages (C4)
inc. English, Polish,
exc. Ancient Greek

Methodology: Base Models

Exploring Large Language Models for Classical Philology

Frederick Riemenschneider
Dept. of Computational Linguistics
Heidelberg University
69120 Heidelberg
riemenschneider@cl.uni-heidelberg.de

Anette Frank
Dept. of Computational Linguistics
Heidelberg University
69120 Heidelberg
frank@cl.uni-heidelberg.de

mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer

Linting Xue* **Noah Constant*** **Adam Roberts***
Mihir Kale **Rami Al-Rfou** **Aditya Siddhant** **Aditya Barua** **Colin Raffel**
Google Research

bowphs/**GreTa**

Pre-trained on: Ancient Greek

bowphs/**PhilTa**

Pre-trained on:
Ancient Greek, Latin, English

-base size

google/**mt5-base**

more parameters = better performance?

google/**mt5-large**

Methodology: Input Encoding

Baseline

Παῦλος <0> καὶ <0> Σιλουανὸς → Text Emb.

Tags Within Text

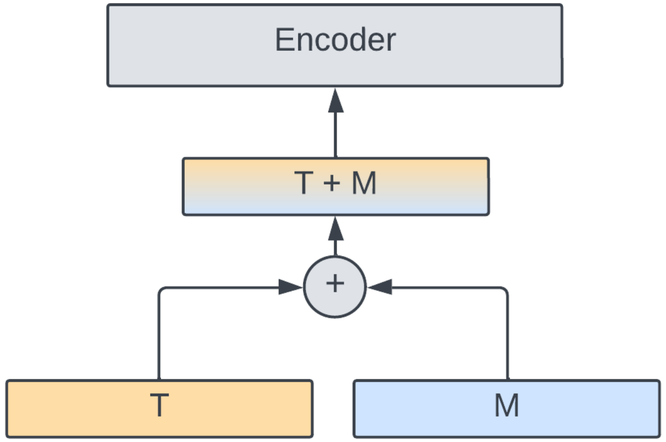
Παῦλος <1> N-NMS <0>
καὶ <1> Conj <0>
Σιλουανὸς <1> N-NMS → Text Emb.

Morphological Embeddings

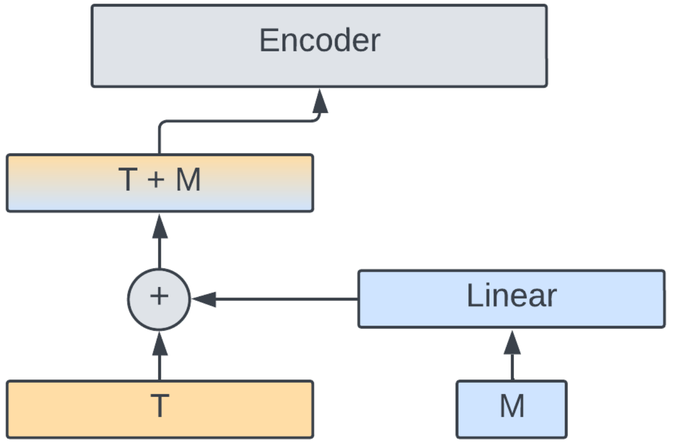
Παῦλος <0> καὶ <0> Σιλουανὸς → Text Emb.

N-NMS <0> Conj <0> N-NMS → Morph. Emb.

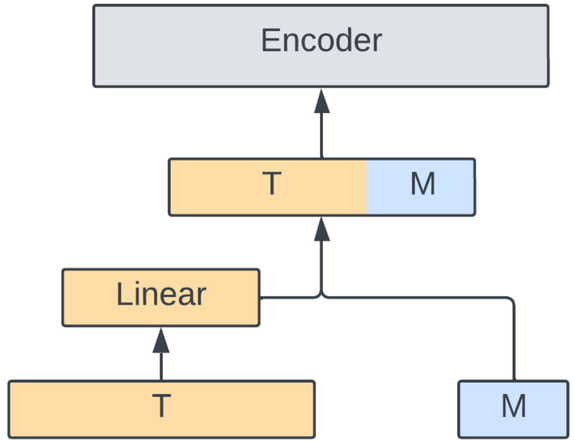
Methodology: Morphological Embeddings



emb-sum

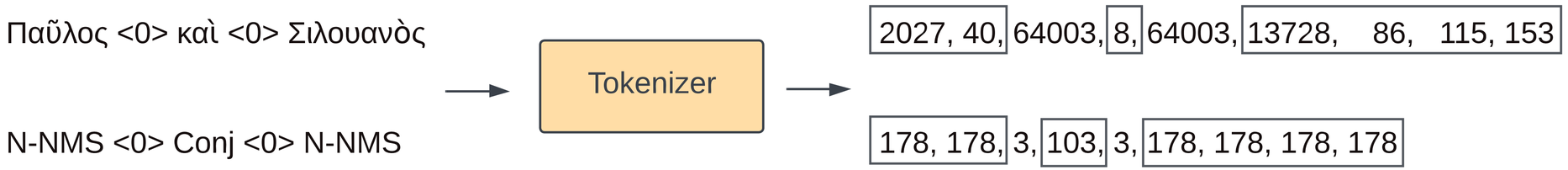


emb-auto



emb-concat

Methodology: Tokenization



Experimental Setups

- 2 target languages (pl, en)
- 4 base models (GreTa, PhilTa, mT5-{base, large})
- 5 input encoding types (baseline, tags-within-text, emb-{sum, auto, concat})
- 2 normalization strategies (raw, normalized)
- 2 tag sets (BibleHub, Oblubienica)

  = 144 unique parameter sets

Ἐγένετο δὲ, ἐν τῷ τὸν Ἀπολλῶ εἶναι ἐν Κορίνθῳ...

εγενετο δε εν τω απολλω ειναι εν κορινθω..

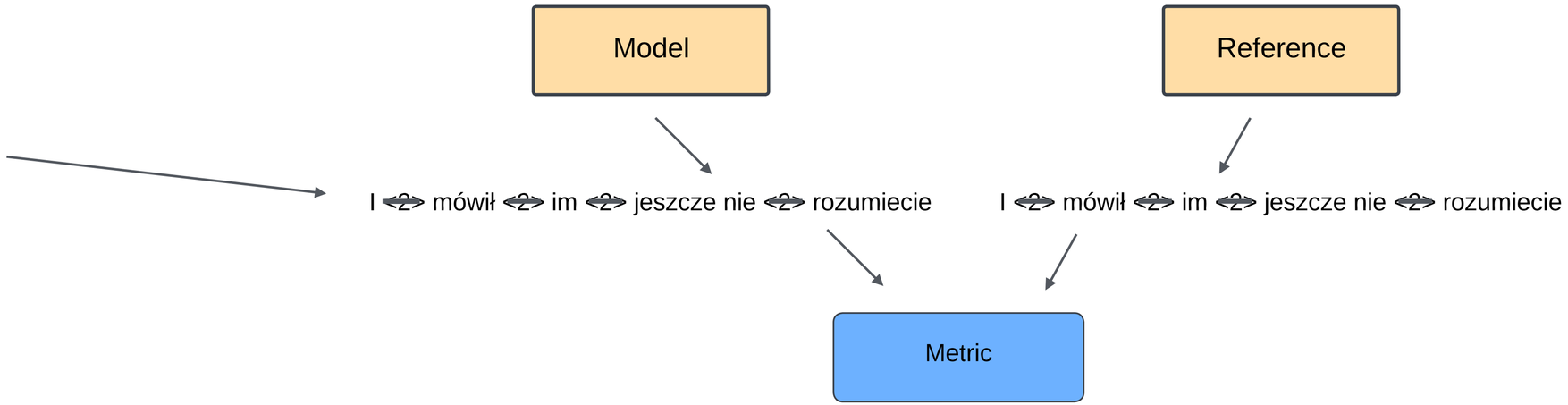
vi Impf Act 3 Sg

V-IIA-3S

Evaluation

- Metrics:
  - BLEU
  - SemScore (all-mpnet-base-v2)

- Fair Evaluation:
  - Skip Separators
  - Trim



Evaluation

- Metrics:
  - BLEU
  - SemScore (all-mpnet-base-v2)

- Fair Evaluation:
  - Skip Separators
  - Trim

(1) παυλος | αποστολος | ~~χριστου | ιη~~
(2) παυλος | αποστολος | ~~χριστου | ιησου | δια | θεληματος | θεου | κατ |~~
(3) παυλος | αποστολος | ~~χριστου | ιησου~~
(4) παυλος | αποστολος | ~~χριστου | ιησου | δι~~
(5) παυλος | αποστολος | ~~χρισ~~
(6) παυλος | αποστολος | ~~χριστου | ιησου | δια | θεληματος | θεου | κατ | επαγγελιαν | ζωης | της | εν | χριστω | ιησου~~

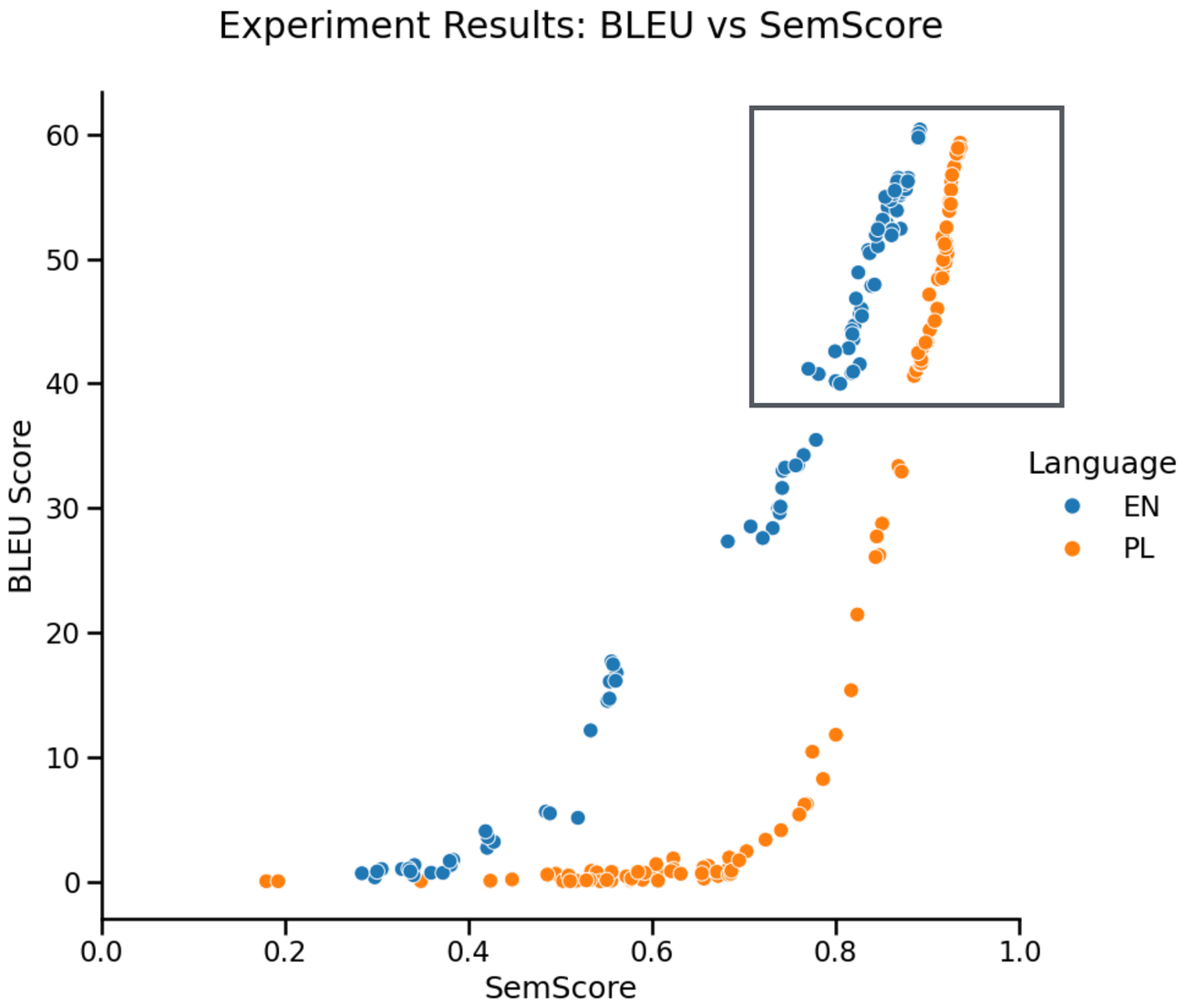
Target:
Paweł | wysłannik | ~~Pomazańca | Jezusa | przez | wolę | Boga | według | obietnicy | życia | - | w | Pomazańcu | Jezusie~~

Results

V-AIA-3S

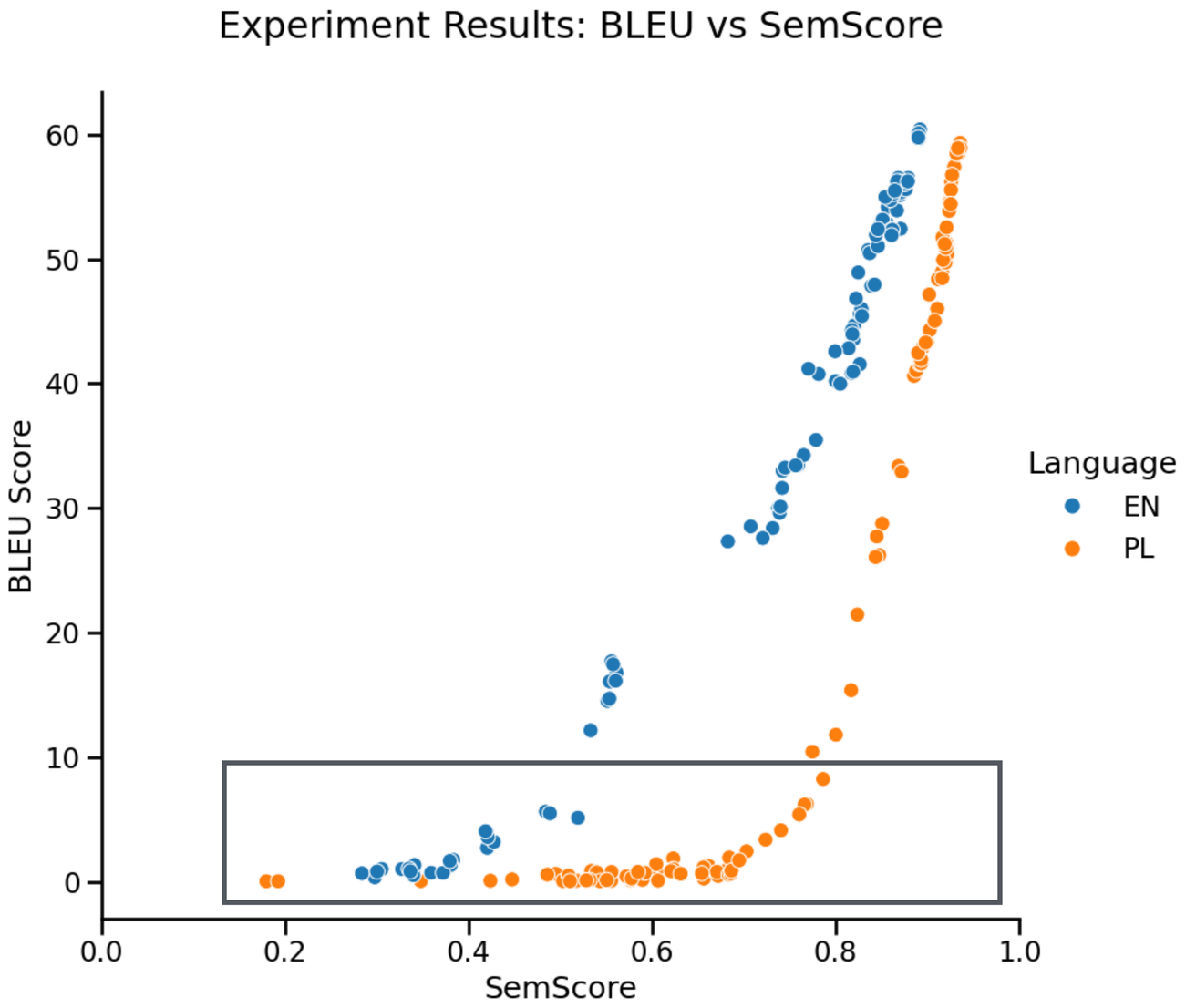
Results: Overall

- How well do the modern Machine Translation models perform on the interlinear translation task?



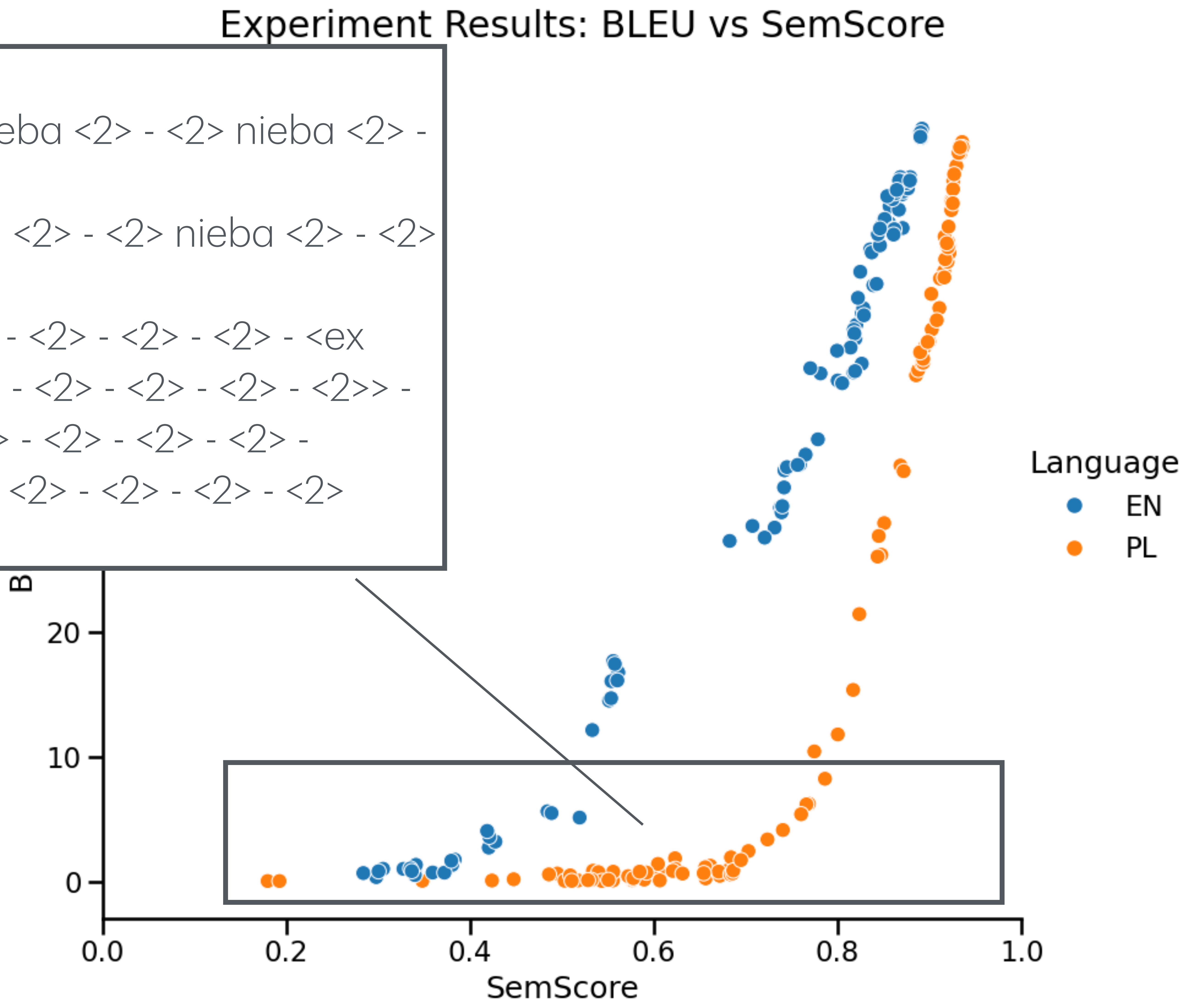
Results: Overall

- How well do the modern Machine Translation models perform on the interlinear translation task?



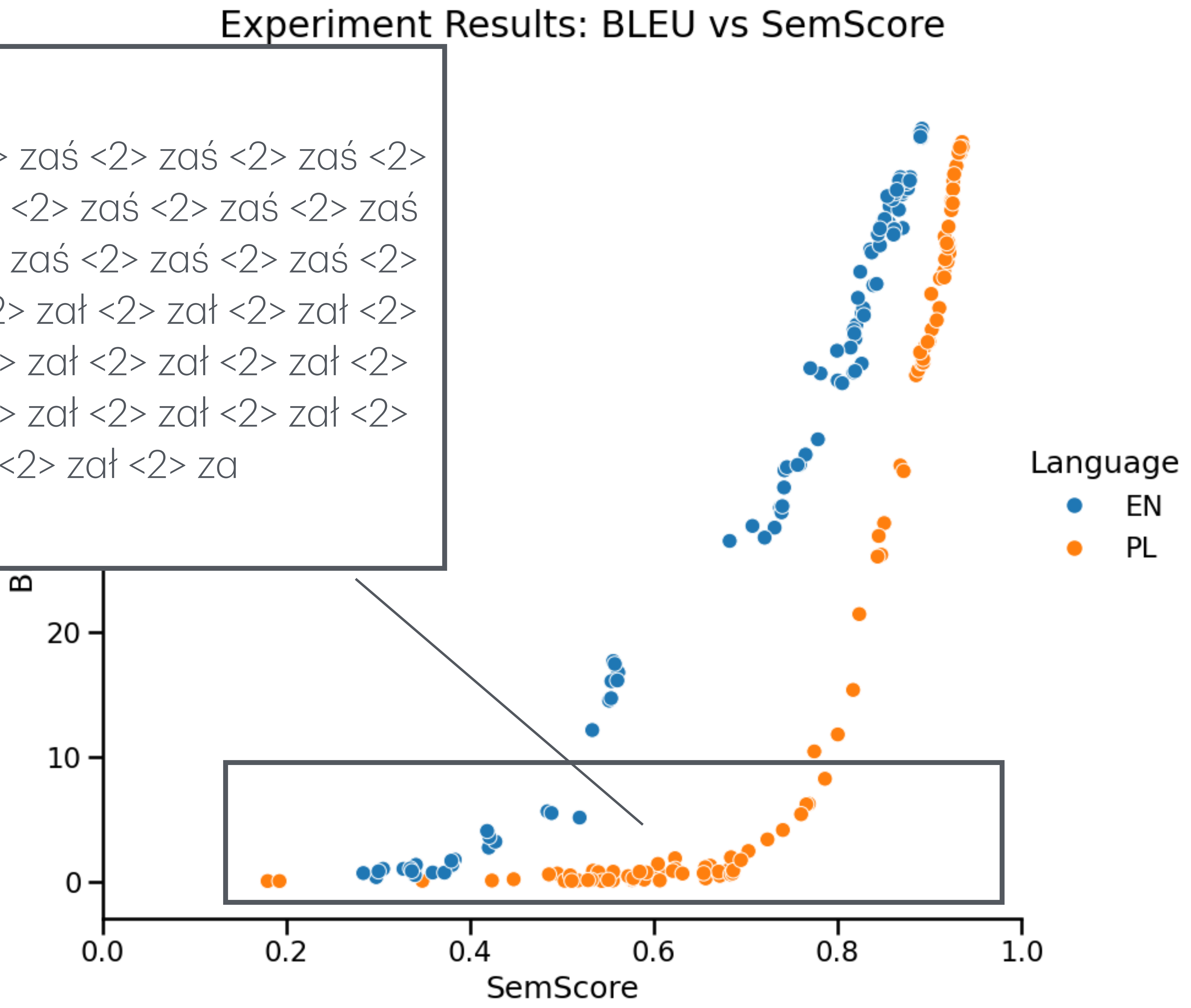
Results: Overall

- <2> zaś <2> - <2> nieba <2> - <2> nieba <2> - <2> nieba <2> - <2> nieba <2> -
<2>
nieba <2> - <2> nieba <2> - <2> nieba <2> - <2> nieba <2> - <2> nieba <2> - <2>
- <2> n
ie <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <ex
tra_id_2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2>> -
<2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> -
<2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2> - <2>



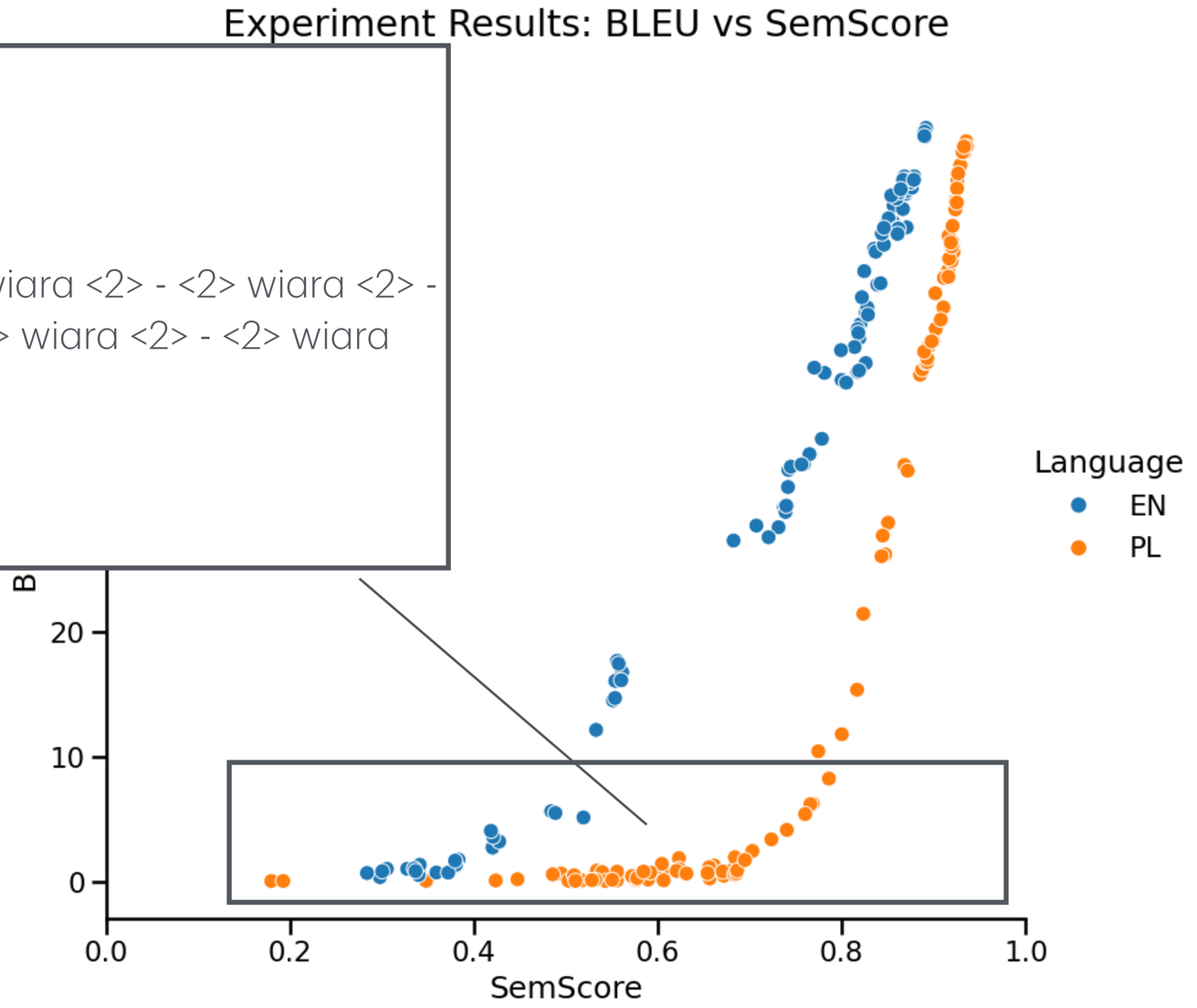
Results: Overall

i <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2>
zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś
<2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2> zaś <2>
zaś <2> zaś <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2>
zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2>
zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2>
zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> zał <2> za



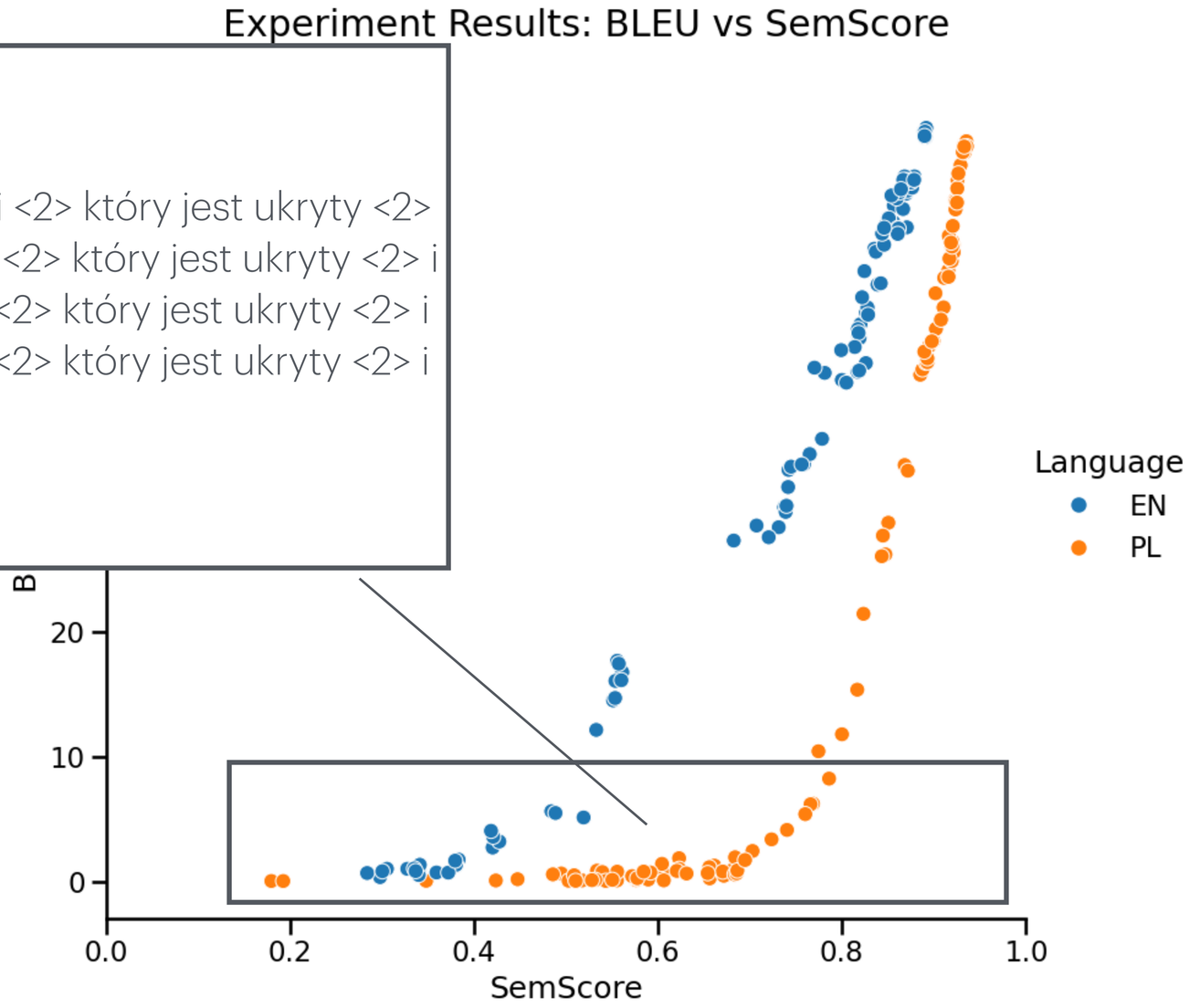
Results: Overall

- <2> zaś <2> - <2> wiara <2> - <2> wiara <2> - <2> wiara <2> - <2> wiara <2> - <2> wiara <2> - <2> wiara <2> - <2> wiara <2> - <2> wiara



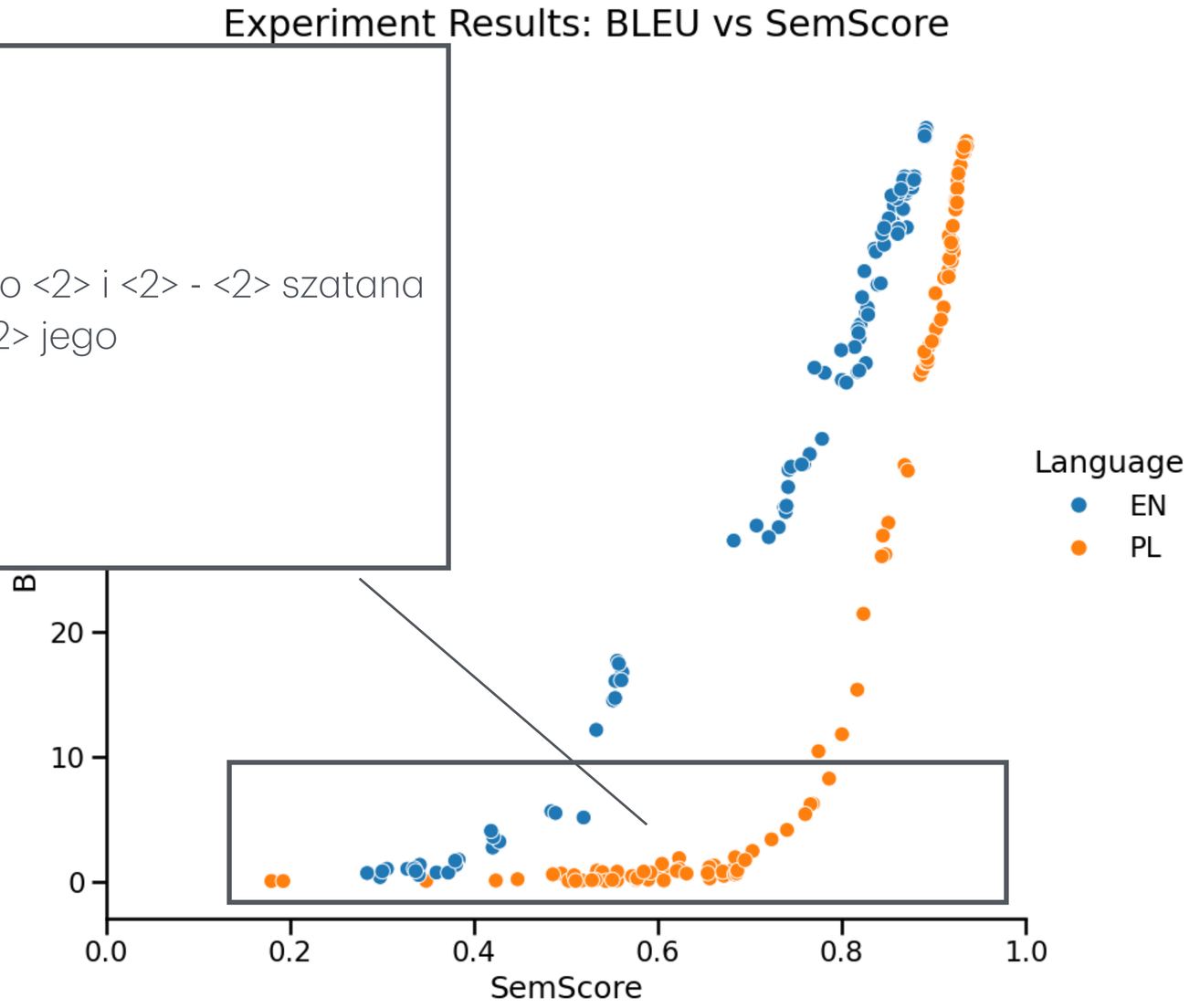
Results: Overall

- <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który jest ukryty <2> i <2> który



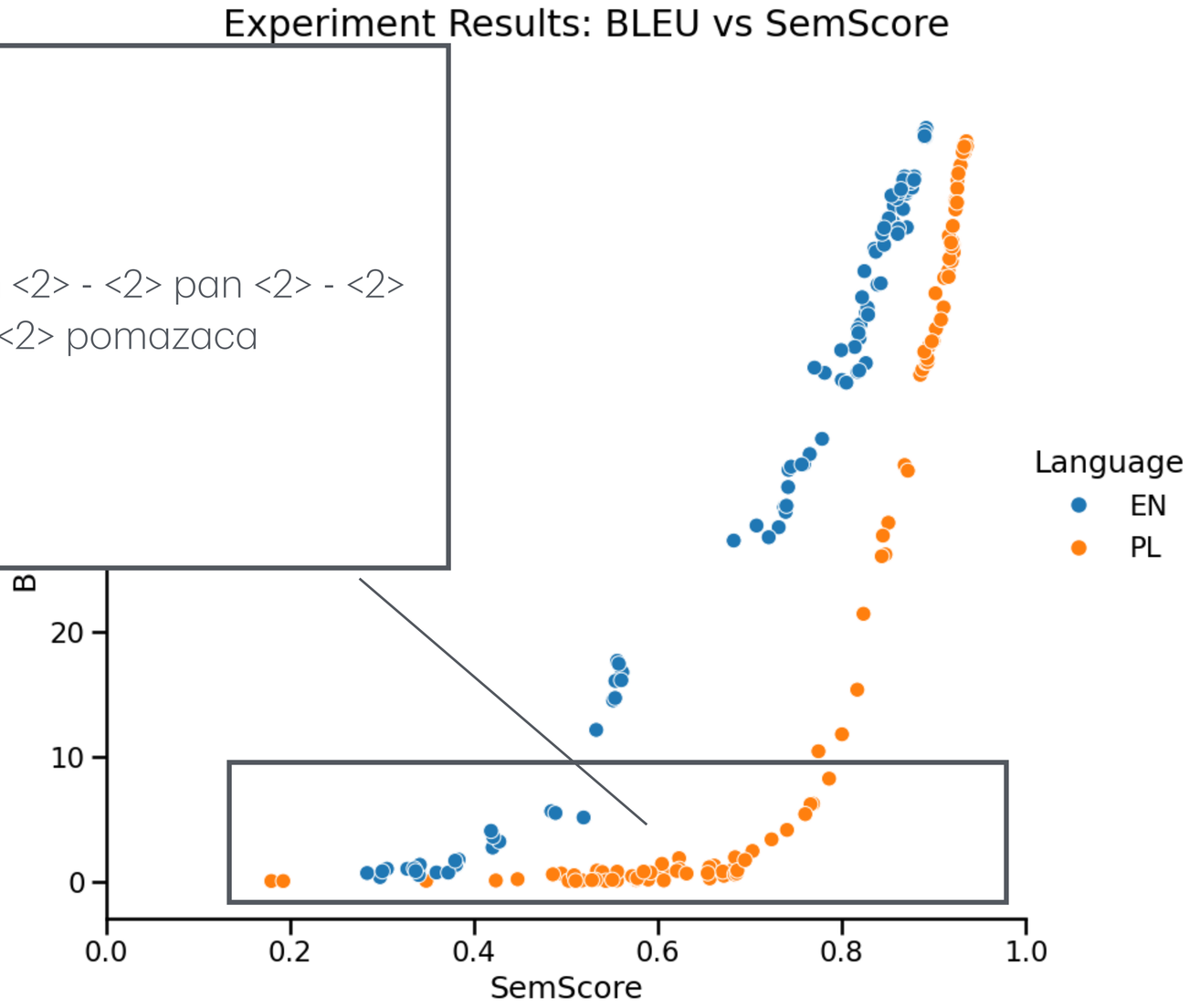
Results: Overall

i <2> zostao dane <2> mu <2> - <2> szatana <2> jego <2> i <2> - <2> szatana <2> jego <2> i <2> - <2> szatana <2> jego



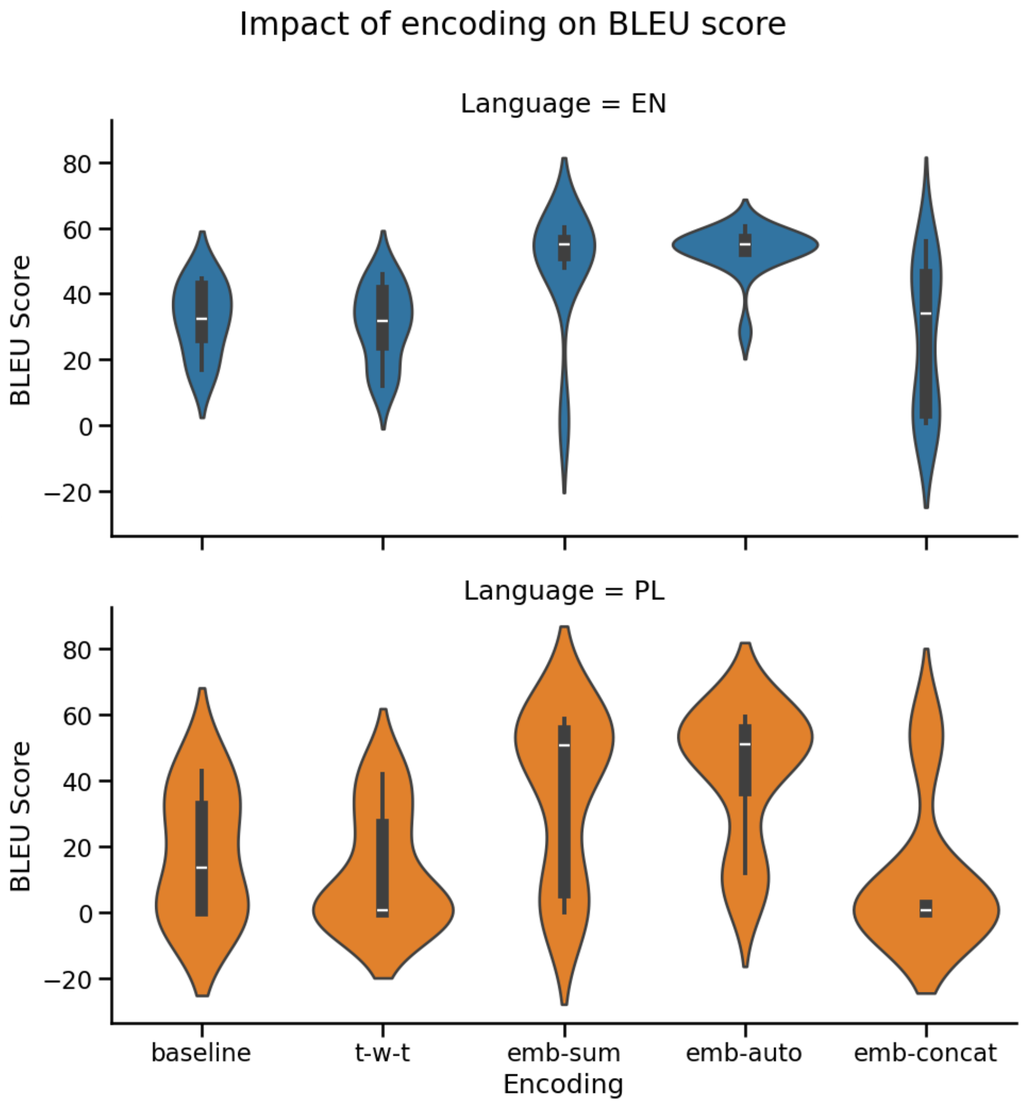
Results: Overall

i <2> powiedzia <2> im <2> - <2> pan <2> - <2> pan <2> - <2> pan <2> - <2> pan <2> i <2> - <2> pomazaca <2> jezusa <2> pomazaca



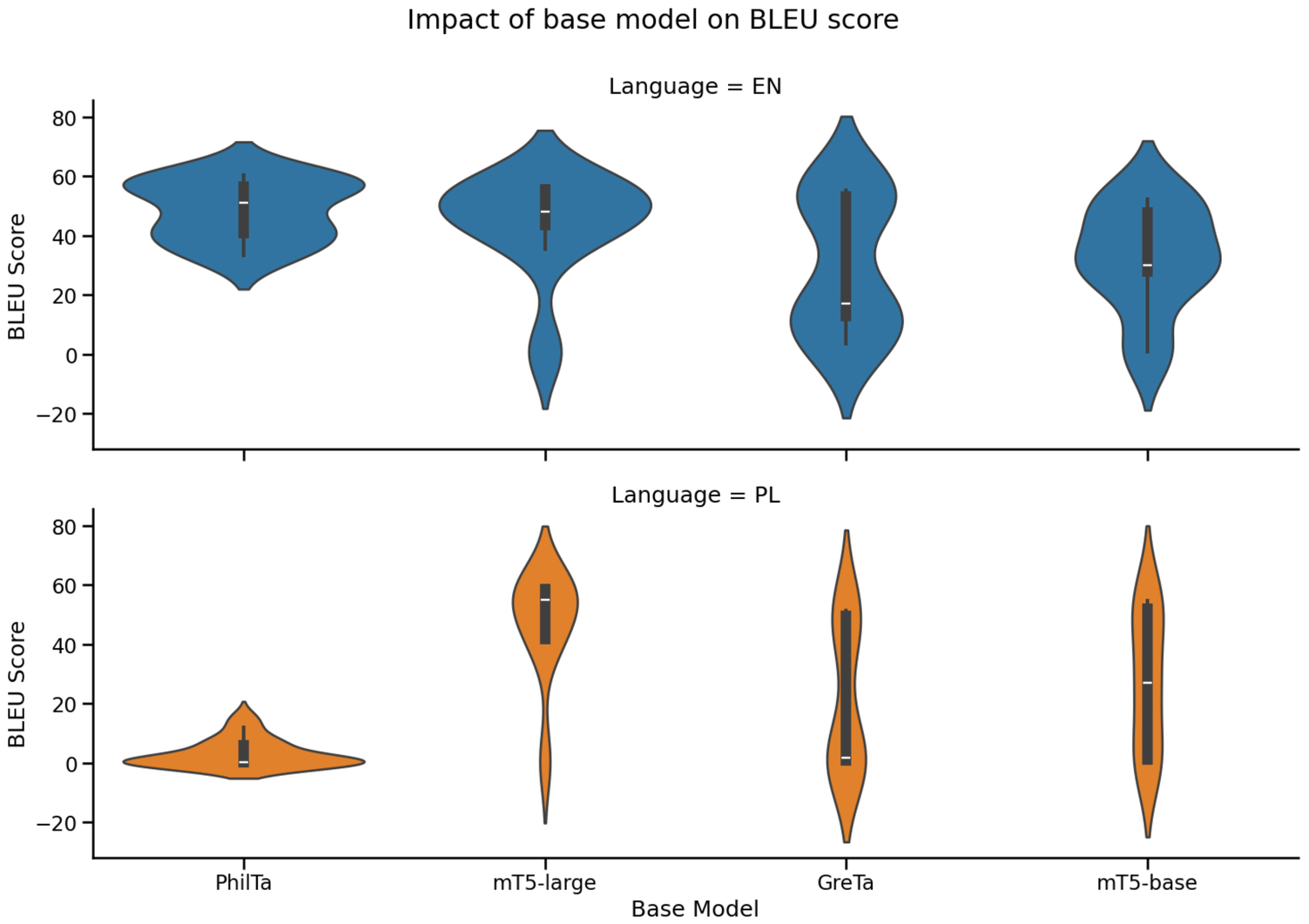
Results: Encodings

- Can morphological tags improve the quality of interlinear translation?



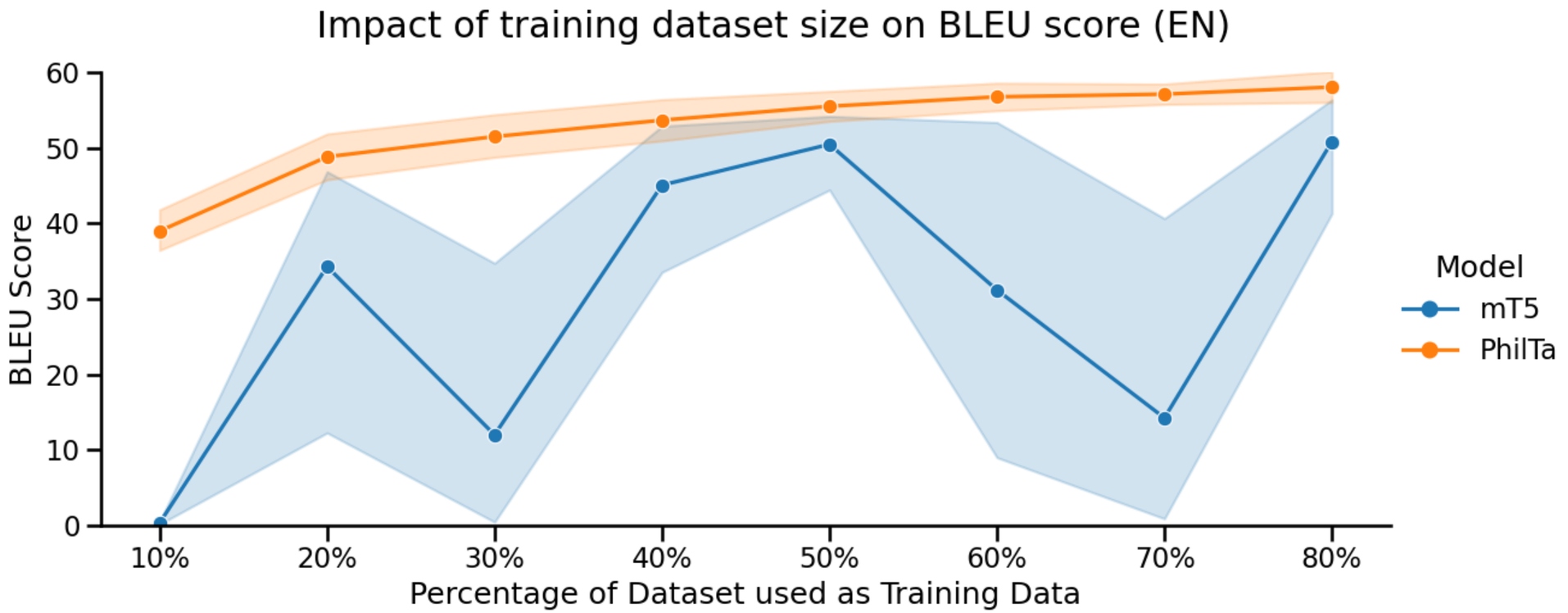
Results: Base Models

- How do specialized ancient language models (PhilTa, GreTa) compare to general multilingual models (mT5)?



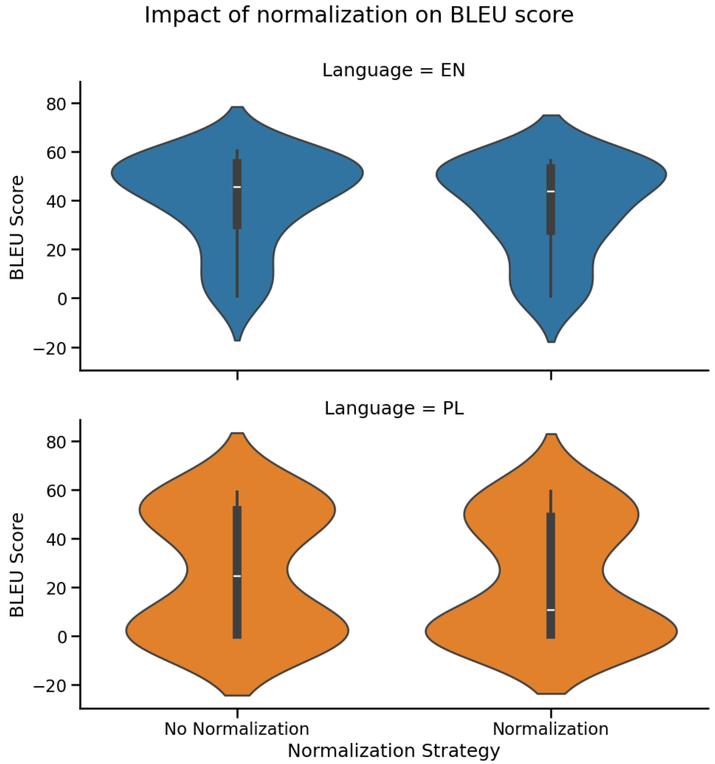
Results: mT5-large vs PhilTa

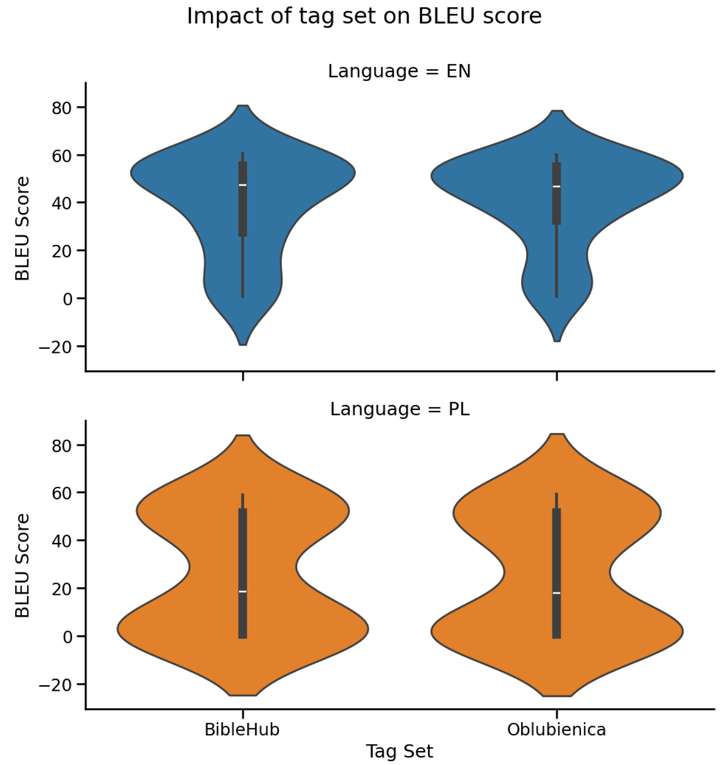
- How much data does {mT5-large, PhilTa} need to achieve sensible performance?



Results: tag sets, preprocessing

- What impact do text preprocessing methods and choice of specific morphological tag sets have on translation performance?





Low-Resource Interlinear Translation: Morphology-Enhanced Neural Models for Ancient Greek

Maciej Rapacz
AGH University of Krakow
mrapacz@agh.edu.pl

Aleksander Smywiński-Pohl
AGH University of Krakow
apohllo@agh.edu.pl

Abstract

Contemporary machine translation systems prioritize fluent, natural-sounding output with flexible word ordering. In contrast, *interlinear translation* maintains the source text's syntactic structure by aligning target language words directly beneath their source counterparts. Despite its importance in classical scholarship, automated approaches to interlinear translation remain understudied.

We evaluated neural interlinear translation from Ancient Greek to English and Polish using four transformer-based models: two Ancient Greek-specialized (GreTa and PhilTa) and two general-purpose multilingual models (mT5-base and mT5-large). Our approach introduces novel morphological embedding layers and evaluates text preprocessing and tag set selection across 144 experimental configurations using a word-aligned parallel corpus of the Greek New Testament.

Results show that morphological features through dedicated embedding layers significantly enhance translation quality, improving BLEU scores by 35% (44.67 → 60.40) for English and 38% (42.92 → 59.33) for Polish compared to baseline models. PhilTa achieves state-of-the-art performance for English, while mT5-large does so for Polish. Notably, PhilTa maintains stable performance using only 10% of training data.

Our findings challenge the assumption that modern neural architectures cannot benefit from explicit morphological annotations. While preprocessing strategies and tag set selection show minimal impact, the substantial gains from morphological embeddings demonstrate their value in low-resource scenarios.[1]

[1]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/2024/017156. The research presented in this paper was partially supported by the funds of Polish Ministry of Science and Higher Education assigned to the AGH University of Kraków.

1 Introduction

Machine translation (MT) is a well-established subfield in Natural Language Processing (NLP), primarily focused on producing accurate and natural translations. In typical scenarios, MT systems have the flexibility to reorder words or go beyond literal meanings to account for syntactic differences between source and target languages. While these conventional MT systems prioritize natural and fluent translations, there exists a spectrum of translation approaches, ranging from free translation to extremely literal renderings.

At the far end of this spectrum lies *interlinear translation* (Shuttleworth and Cowie, 2014), a method that strictly preserves the source text's syntactic structure. This approach aligns target language words directly below or above their corresponding source text elements. Commonly applied to ancient (and oftentimes sacred) texts, this method allows readers unfamiliar with the source language to understand both the meaning and structure of the original text. Such alignment enables students to critically evaluate translations by observing how specific source words were translated, which is especially crucial for interpreting source texts in fields such as philosophy and religious studies. Figure 1 illustrates an example of interlinear translation.

Despite the significance of interlinear translation, which Benjamin (1923/2000) called "the archetype or ideal of all translation", there has been limited research on automating this process. This may be attributed to the pre-existing interlinear translations for many influential texts. However, we believe automating this process remains relevant, making these texts more accessible to those without expertise in ancient languages.

While prior research (Tenney et al., 2019) suggests that modern neural architectures like BERT inherently learn linguistic patterns without explicit

Proceedings of the First Workshop on Language Models for Low-Resource Languages (LoResLM 2025), pages 145–165
January 20, 2025.