TOWARDS A METRIC FOR PARAPHRASTIC MODIFICATION

ABSTRACT AVAILABLE AT: HTTPS://DH2018.ADHO.ORG/EN/TOWARDS-A-METRIC-FOR-PARAPHRASTIC-MODIFICATION/

Maria Moritz,[1] Johannes Hellrich,[2,3] and Sven Büchel[3]

[1] eTRAP Research Group, University of Goettingen, Germany
[2] Graduate School "The Romantic Model", Friedrich-Schiller-Universität Jena, Germany
[3] JULIE Lab, Friedrich-Schiller-Universität Jena, Germany

Digital Humanities 2018, Mexico City, June 26-29

eTRAP
Electronic Text Reuse Acquisition Project
GÖTTINGEN CENTRE FOR DIGITAL HUMANITIES

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

SPONSORED BY THE
Federal Ministry of Education and Research

MODIFICATIONS IN PARAPHRASTIC REUSE

Text Reuse is the written repetition of text. In historical texts, text reuse detection challenges include:

language variation over time

incomplete witnesses

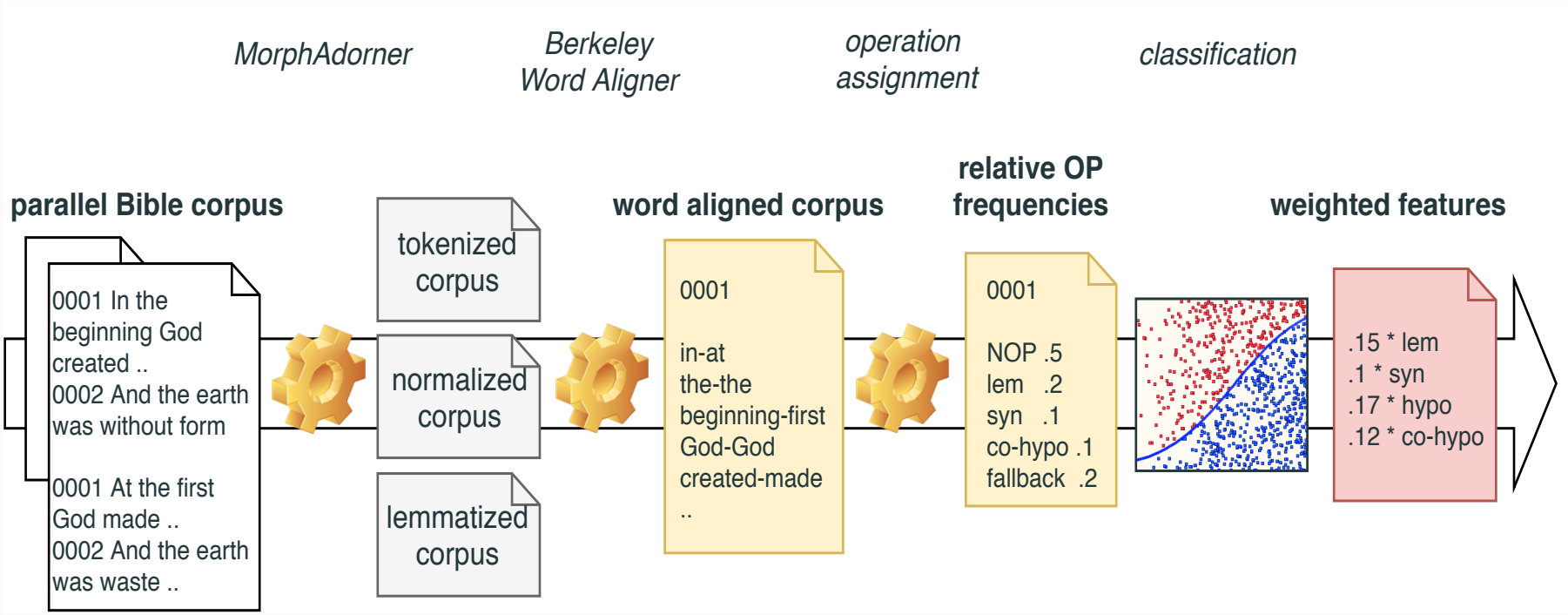
diverse reuse types

To further improve automatic detection of paraphrase, we:

quantify paraphrastic modification

to better understand reuse

METHOD


MorphAdorner
Berkeley Word Aligner
operation assignment
classification
parallel Bible corpus
0001 In the beginning God created ..
0002 And the earth was without form
0001 At the first God made ..
0002 And the earth was waste ..
tokenized corpus
normalized corpus
lemmatized corpus
word aligned corpus
0001
in-at
the-the
beginning-first
God-God
created-made
..
relative OP frequencies
0001
NOP .5
lem .2
syn .1
co-hypo .1
fallback .2
weighted features
.15 * lem
.1 * syn
.17 * hypo
.12 * co-hypo

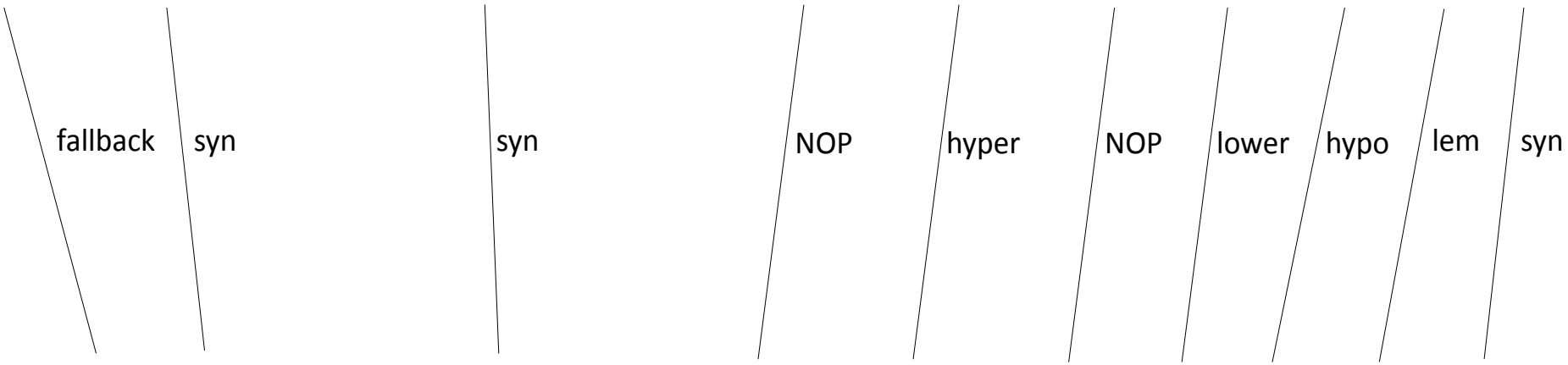
PARALLEL BIBLE CORPUS AS TEXT DATA

Name	Abbr.	Publication	Source	Translation Class
The Webster Bible	WBT	1833	bst[1]	standard
English Revised Version	ERV	1881-1894	mys[2]	standard
Darby Bible	DBY	1890	ptp[3]	standard
Brenton's English Septuagint	*LXXE*	*1844-1884*	*mys*[2]	*literal*
Young's Literal Translation	*YLT*	*1862*	*bst*[1]	*literal*
Smith's Literal Translation	*SLT*	*1876*	*mys*[2]	*literal*

[1] `http://www.biblestudytools.com/`
[2] `www.mysword.info/`
[3] Mayer & Cysouw 2014

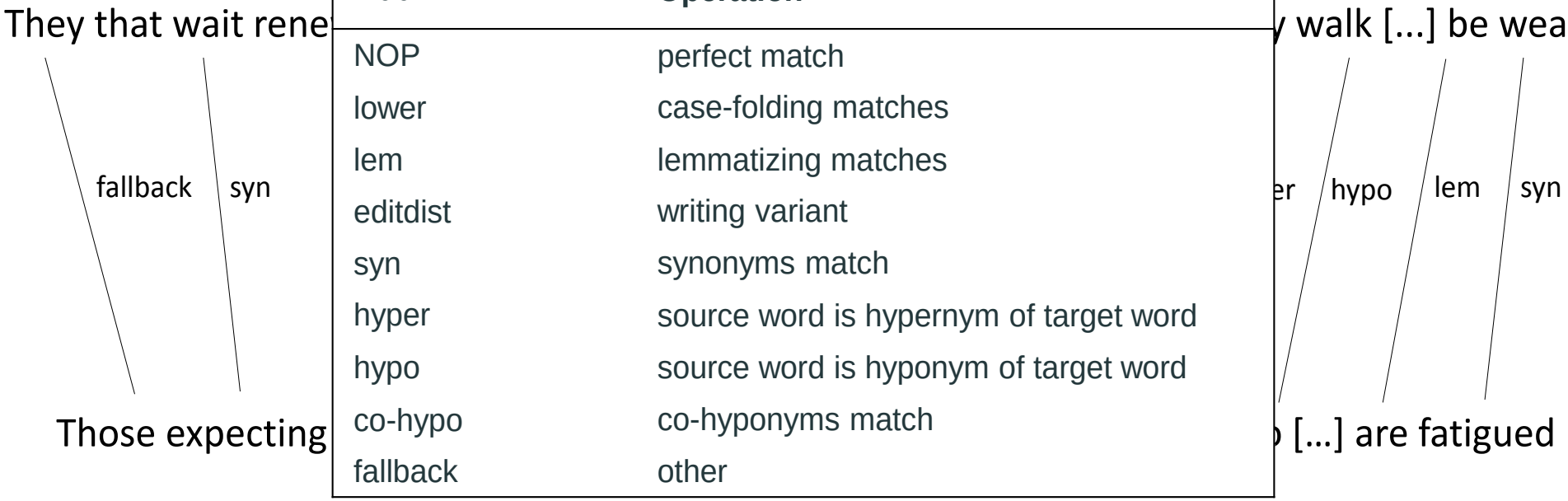
ALIGNMENT AND OPERATION ASSIGNMENT

They that wait renew their strength they mount up with wings as eagles they walk [...] be weary



Those expecting pass to power They raise up the pinion as eagles They go […] are fatigued

ALIGNMENT AND OPERATION ASSIGNMENT



Abbr.	Operation
NOP	perfect match
lower	case-folding matches
lem	lemmatizing matches
editdist	writing variant
syn	synonyms match
hyper	source word is hypernym of target word
hypo	source word is hyponym of target word
co-hypo	co-hyponyms match
fallback	other

OPERATION RATIOS FOR INSPECTION

Darby's Bible Translation & the English Revised Version

id	NOP	lower	lem	editdist	syn	hyper	hypo	co-hypo	fallback
01001001	0.889	0.0	0.111	0.0	0.0	0.0	0.0	0.0	0.0
01001002	0.815	0.037	0.0	0.0	0.037	0.0	0.0	0.0	0.111
01001003	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

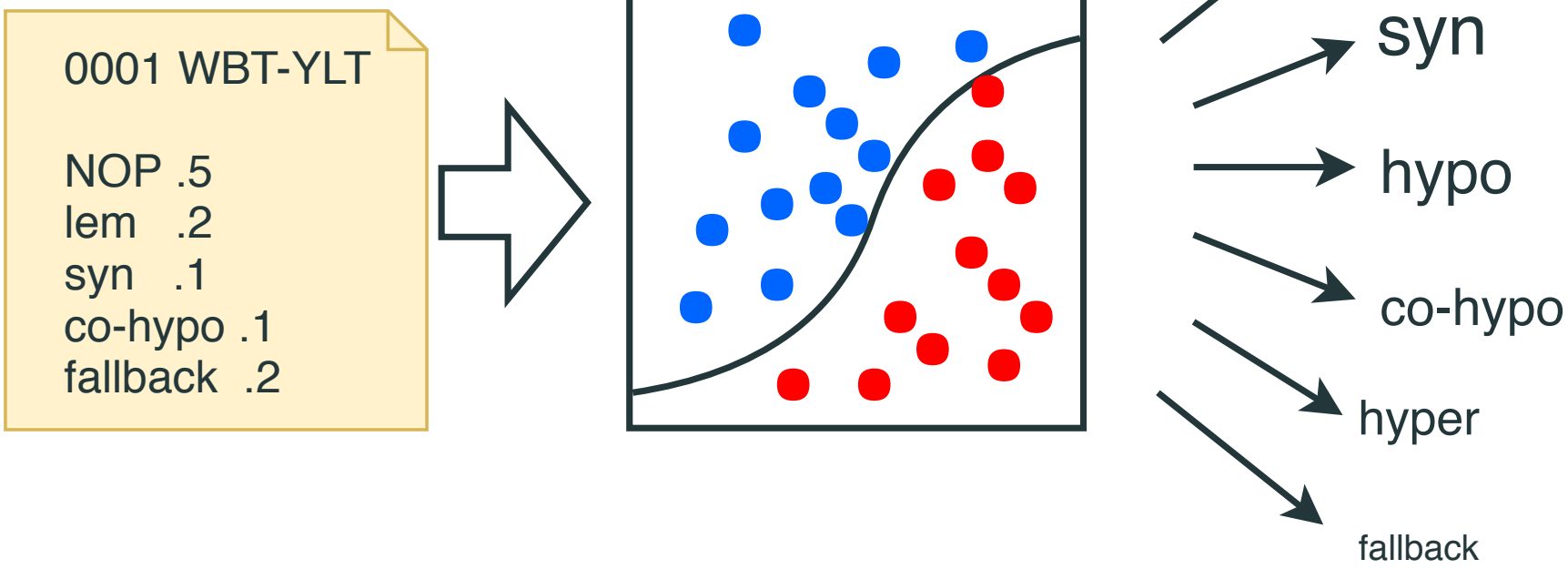
Young's & Smith's Literal Translation

id	NOP	lower	lem	editdist	syn	hyper	hypo	co-hypo	fallback
01001001	0.778	0.0	0.111	0.0	0.0	0.0	0.0	0.111	0.0
01001002	0.72	0.04	0.0	0.0	0.04	0.04	0.0	0.0	0.16
01001003	0.714	0.0	0.143	0.0	0.0	0.0	0.0	0.0	0.143

LOGISTIC REGRESSION MODELING (YES/NO-DECISION)

relative OP frequencies = features

aligned Bibles from same/different class?



FEATURE WEIGHTS

Label(x_i)	Operation	Estimated coefficient θ_{relative}
NOP	perfect match	dropped
lower	case-folding matches	0.060
lem	lemmatizing matches	0.195
low_editdist	writing variant	0.068
syn	synonyms match	0.190
hyper	source word is hypernym of target word	0.117
hypo	source word is hyponym of target word	0.170
co-hypo	co-hyponyms match	0.122
fallback	other	0.078
		= 1.000

FEATURE WEIGHTS

Label(x_i)	Operation	Estimated coefficient $\theta_{relative}$
NOP	perfect match	dropped
lower	case-folding matches	0.060
lem	lemmatizing matches	0.195
low_editdist	writing variant	0.068
syn	synonyms match	0.190
hyper	source word is hypernym of target word	0.117
hypo	source word is hyponym of target word	0.170
co-hypo	co-hyponyms match	0.122
fallback	other	0.078
		= 1.000

OUR SCORE CONSIDERS WEIGHTED FEATURES

Based on the estimated coefficients, and the operation frequencies, we define the paraphrasticality par between two word-aligned text spans a and b as:

$$par(a, b) = \sum_{i=0}^{n} \theta_i x_i^{a,b}$$

VALIDATION OF PAR(A,B)

		b				
		ERV	WBT	*LXXE*	*YLT*	*SLT*
	DBY	0.13	0.13	0.29	0.31	0.29
	ERV	-	0.09	0.3	0.32	0.31
a	WBT	-	-	0.28	0.33	0.29
	LXXE	-	-	-	0.42	0.37
	YLT	-	-	-	-	0.31

similar — distant

DBY, ERV, WBT = standard translations

LXXE, YLT, SLT = literal translations

SUMMARY

We present a technique to measure paraphrasticality that:

- describes paraphrasticality as frequency of modifications for which we find empirical weights, and
- is specifically useful for applications in the humanities as operation frequencies and weights are open to manual inspection.
- Future Work: a comprehensive comparison against existing metrics (e.g., BLEU, METEOR) will be published (Moritz et al. 2018, LaTeCH).

Thank you!

Presentation

Maria Moritz

https://www.etrap.eu mmoritz@etrap.eu

SPONSORED BY THE

Federal Ministry of Education and Research

Co-authors

Johannes Hellrich

http://modellromantik.uni-jena.de johannes.hellrich@uni-jena.de

Sven Büchel

http://julielab.de sven.buechel@uni-jena.de

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

Jena University Language & Information Engineering Lab

MODELL ROMANTIK
Variation • Reichweite • Aktualität

DFG Deutsche Forschungsgemeinschaft

APPENDIX

REFERENCES

- Brill, Eric and Moore, Robert C. (2000): An Improved Error Model for Noisy Channel Spelling Correction. Proceedings of ACL 2000.
- Burns, Philip R (2013): Morphadorner v2: A java library for the morphological adornment of English language texts. `http://morphadorner.northwestern.edu`.
- Daniell, David (2003): The Bible in English: its history and influence. New Haven, Conn: Yale University Press.
- De Nero, J and Klein, D (2007): Tailoring word alignments to syntactic machine translation. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. Prague, pp. 17–24.
- Denkowski, M., & Lavie, A. (2014): Meteor universal: Language specific translation evaluation for any target language. In: *Proceedings of the ninth workshop on statistical machine translation*, pp. 376-380.
- Finch, Andrew, Hwang, Young-Sook, & Sumita, Eiichiro (2005): Using machine translation evaluation techniques to determine sentence-level semantic equivalence. In: *Proceedings of the Third International Workshop on Paraphrasing (IWP2005)*.
- Li, Ming and Vitáni, Paul (2008): An Introduction to Kolmogorov Complexity and Its Applications 3. Auflage. Springer New York.
- Malone, David (2010): Julia Smith bible translation (1876). Online: `https://recollections.wheaton.edu/2010/12/julia-smith-bible-translation-1876/`.
- Marlowe, Michael (2005): Webster's Revision of the KJV (1833). Online: `http://www.bible-researcher.com/webster.html`.
- Marlowe, Michael (2017): John Nelson Darby's Version. Online: `http://www.bibleresearcher.com/darby.html`.
- Mayer, Thomas & Cysouw, Michael (2014): Creating a massively parallel bible corpus. In: *Proceedings of LREC'14*. European Language Resources Association (ELRA).
- Moritz, Maria, Hellrich, Johannes & Büchel, Sven (forthcoming 2018): A Human-Interpretable Method to Predict Paraphrasticality, accepted for LaTeCH-CLfL *@* COLING.
- Navigli, Roberto, and Ponzetto, Simone Paolo (2012): Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.*
- Newman, Card. John Henry (1859): The text of the rheims and douay version of holy scripture. The Rambler, 1(New Series, Part II). 145–169.
- Pope, Hugh (1910): The origin of the douay bible. The Dublin Review, 147(294-295), pp. 97–118.
- Roger, Nicole (1958): New Testament Use of the Old Testament. In: *Henry, C.F.H. (ed.), Revelation and the Bible. Contemporary Evangelical Thought*. Grand Rapids: Baker, 1958. London: The Tyndale Press, 1959.
- Su, Keh-Yih, Wu, Ming-Wen and Chang, Jing-Shin (1992): A New Quantitative Quality Measure for Machine Translation Systems. In: *Proceedings of the 14th conference on Computational linguistics*, pp. 433-439. ACL.
- Young, Robert (1898a): Young's Translation: Publisher's Note and Preface. Online: `http://www.ccel.org/bible/ylt/ylt.htm`
- Young, Robert (1898b): Young's Literal Translation. Online: `http://www.bibleresearcher.com/young.html`

LICENCE