TEXT REUSE IN BERNARD DE CLAIRVAUX:

HOW DO LEXICAL RESOURCES HELP MODEL TEXT REUSE?

Maria Moritz and Marco Büchler

August 6-8 2017, *Digital Humanities 2017, Montréal*

eTRAP
Electronic Text Reuse Acquisition Project

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

TABLE OF CONTENTS

INTRODUCTION

Text Reuse:

- Spoken and written repetition of text across time and space
- E.g., citations, allusions, and translations

Text is often modified during the reuse process:

	Bible verse	Bernard reuse
Psalm 20 3	desiderium animae eius tribuisti ei et voluntate labiorum eius non fraudasti eum diapsalma	desiderium cordis eius tribuetur ei ,
English	Remember all thy offerings, and accept thy burnt sacrifice	his heart's desire will be given to [him]

Long-term goal of Reuse Style Analysis:

- A formal model of reuse transformations,
- Empirical insight into the amount and type of reuse modifications,

Identify morphological changes & word substitutions

- Using Ancient Greek WordNet & LWN (AGWN, Bizzoni et al., 2014, Latin WordNet, Minozzi, 2009) & BabelNet (BN, Navigli & Ponzetto, 2012)
- In excerpts from sermons, treatises, letters from the Sources Chrétiennes collection
- Extracted by the Biblindex project (Mellerin, 2014), each pointing to a Bible verse
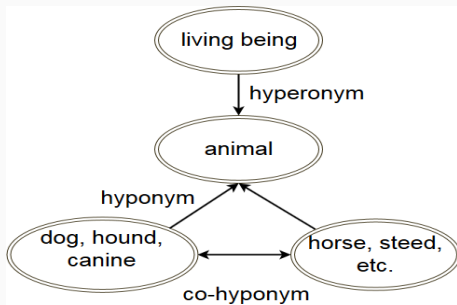- 1,128 unique reuse-to-bible-verse pairs

METHODOLOGY

LINGUISTIC RESOURCES
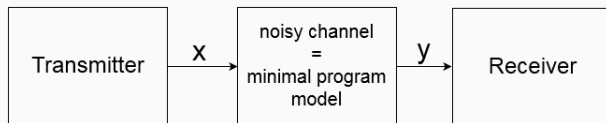
For word modifications:

- Biblindex' lemma list: 315,021 entries
- Classical Language Tool Kit (CLTK, Johnson et al. '14): 270,228 entries

For word substitutions:

- AGWN (LWN): 99K synsets, 27K contain Latin words
- BabelNet: lang: 271; synsets: >13 Mio. (1.5K for Bernard); senses: >745 Mio.

To model the transformation in-between two text excerpts, we introduce operations.[1]

operation	example
NOP(reuse_word, orig_word)	*NOP(maledictus,maledictus)*
upper(reuse_word, orig_word)	*upper(filio,Filio)*
lower(reuse_word, orig_word)	*lower(Gloriam, gloriam)*
lem(reuse_word, orig_word)	*lem(penetrat, penetrabit)*
repl_syn(reuse_word, orig_word)	*repl_syn(magnificavit, glorificavit)*
repl_hyper(reuse_word, orig_word)	*hyper(cupit, habens)*
repl_hypo(reuse_word, orig_word)	*hypo(dederit, tollet)*
repl_co-hypo(reuse_word, orig_word)	*repl_co-hypo(magnificavit, fecit)*
lemma_missing(reuse_word, orig_word)	*lemma_missing(tentari, inlectus)*
no_rel_found(reuse_word, orig_word)	*no_rel_found(gloria, arguitur)*

An algorithm identifies those operations in a prioritized order.

[1]Moritz et al. 2016. Non-literal text reuse in historical texts: An approach to identify reuse transformations and its application to bible reuse. *EMNLP'16*.

APPROACH – ALGORITHM DESIGN

- Find all possible OPs for a reuse word from the counter verse
- Apply the most literal operation and blacklist the counter word
- Proceed with the next word from the reuse

$R \leftarrow$ set of words from the reuse
$B \leftarrow$ set of words from the Bible verse
$BL \leftarrow$ blacklist
$program \leftarrow$ final operation list

1: $BL = \{\emptyset\}$
2: **for** $i = 1$ to R **do**
3:     $operations = \{\emptyset\}$
4:     **for** $j = 1$ to B **do**
5:        $operations = operations \cup op(i, j)$
6:     **end for**
7:     $program = program \cup min(cost(operations\_tmp))$
8:     $BL = BL \cup index(min(cost(operations\_tmp)))$
9: **end for**

We apply the algorithm on Bernard's reuse, first using AGWN, then BN.

RESULTS

A support value determines the ratio of identified non-literal OPs.

- success = {syn, repl_hyper, repl_hypo, repl_co-hypo}
- $sup = \dfrac{\sum_{Occ(o)} o \in success}{\sum_{Occ(o)} o \in success \cup \{no\_rel\_found\}}$
- sup_{AGWN} is about 28%, sup_{BN} is about 3%.

## We group the identified operations by their literalness.

		literal		lem	syn	non-literal[2] hyper	hypo	co-hypo	no_rel_found	unclassified lem_missing	total
	NOP	upper	lower								
AGWN	4521	1	396	770	397	125	124	316	2470	450	9570
BN	4526	1	397	771	25	22	35	27	3316	450	9570
		60%				10%				30%	

Using AGWN, we encounter a high ratio of syn- & co-hyponyms, and a significant no. of hyper- & hyponyms. While using BN these figures are about 1/10.

---

[2]Figures updated.

AGWN outperforms BN in identifying semantic relations ("non-literal")
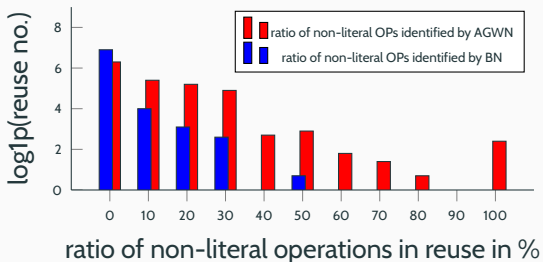
We find 3 patterns:

1. The more semantic related words replaced, the more likely it is an allusion or analogy,
2. Short allusions better covered than paraphrases with many semant. related words,
3. Highly literal reuse is covered best.

Summarizing AGWN better identifies allusions (non-literal reuse).

non-literal ratio	analogy	allusion	paraphrased	near literal
100	6	1	1	1
60 - 80	5	1	1	0
40 - 50	1	1	1	2
10 - 30	0	0	0	5

Table 1: Sample classification[3]

Isaïe 9 6 analogy

Bible: parvulus enim natus est nobis filius datus est nobis et factus est principatus super umerum eius et vocabitur nomen eius Admirabilis consiliarius Deus fortis Pater futuri saeculi Princeps pacis

(a child is born to us a son is given and the government will be on his shoulder and he will be called Wonderful Counselor, Mighty God, Everlasting Father, Prince of Peace)

Reuse: magni consilii Angelus *(great counsel angel)*

OPs: 0:23 syn(magni,fortis); 1:18 hypo(consilii,nomen); 2:22 co-hypo(angelus,deus);

---

[3]Classification by L. Mellerin

CONCLUSION

CONCLUSION

Summary:

- We analyzed the ratio of non-literal reuse in a Latin dataset showing the support of 2 lexical resources.
- Results show that language resources for Latin are limited and support coverage only to a certain extent, which raises awareness for the lack of resources for ancient data.

Future Work:

- Refine the operation set and the algorithm,
- Analyze more languages,
- Increase the size of the datasets,
- Consider the use of word embedding, e.g., to address POS-changes.

REFERENCES

- Yuri Bizzoni, Federico Boschetti, Harry Diakoff, Riccardo Del Gratta, Monica Monachini, and Gregory Crane. 2014. The making of ancient greek wordnet. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).

- Kyle P. Johnson, Patrick J. Burns, Luke Hollis, Martín Pozzi, Amit Shilo, Stephen Margheim, Gitter Badger, and Eamonn Bell. 2014–2016. Cltk: The classical language toolkit. `https://github.com/cltk/cltk`.

- Laurence Mellerin. 2014. New ways of searching with biblindex, the online index of biblical quotations in early christian literature. In Claire Clivaz, Andrew Gregory, and David Hamidovic, editors, *Digital Humanities in Biblical, Early Jewish and Early Christian Studies*, chapter 11, pages 175–192. Brill, Leiden.

- Stefano Minozzi, 2009. *Innsbrucker Beiträge zur Sprachwissenschaft*, volume 137, chapter The Latin WordNet Project, pages 707–716. Institut für Sprachen und Literaturen der Universität Innsbruck, Innsbruck.

- Maria Moritz, Andreas Wiederhold, Barbara Pavlek, Yuri Bizzoni, and Marco Büchler. 2016. Non-literal text reuse in historical texts: An approach to identify reuse transformations and its application to bible reuse. In *Empirical Methods in Natural Language Processing (EMNLP'16), Austin, TX, USA*. Association for Computational Linguistics.

- Roberto Navigli and Simone Paolo Ponzetto. 2012. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.*, 193:217–250, December.

Visit us

🌐 http://www.etrap.eu

✉ contact@etrap.eu

SPONSORED BY THE

eTRAP

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Federal Ministry
of Education
and Research

APPENDIX

- Gribomont J. Weber R., Fischer B., editor. 1969, 1994, 2007. Biblia sacra juxta vulgatam versionem. Deutsche Bibelgesellschaft.

- Bernard de Clairvaux, Sermons sur le Cantique 1-15, Sermones super Cantica Canticorum 1-15, P. A. O à Sources Chrétiennes, col. 414, 1996.
- Bernard de Clairvaux, Sermons sur le Cantique 16-32, Sermones super Cantica Canticorum 16-32 , P. A. O à Sources Chrétiennes, col. 431, 1998.
- Bernard de Clairvaux, Sermons sur le Cantique 33-50, Sermones super Cantica Canticorum 33-50, , P. A. O à Sources Chrétiennes, col. 452, 2000.
- Bernard de Clairvaux, Sermons sur le Cantique 51-68, Sermones super Cantica Canticorum 151-68 , P. A. O à Sources Chrétiennes, col. 472, 2003.
- Bernard de Clairvaux, Sermons sur le Cantique 69-86, Sermones super Cantica Canticorum 69-86 , P. A. O à Sources Chrétiennes, col. 511, 2007.
- Bernard de Clairvaux, Amour de Dieu, Liber de diligendo Deo, P. A. O à Sources Chrétiennes, col. 393, 1993.
- Bernard de Clairvaux, Lettres 1-41, Epistolae 1-41, P. A. O à Sources Chrétiennes, col. 425, 1997.
- Bernard de Clairvaux, Lettres 42-91, Epistolae 42-91, P. A. O à Sources Chrétiennes, col. 425, 1997.
- Bernard de Clairvaux, Lettres 92-163, Epistolae 92-163, P. A. O à Sources Chrétiennes, col. 425, 1997.
- Bernard de Clairvaux, Lettres 363-495, Epistolae 363-495, SBO VIII, 311-447 1957-1977.
- Bernard de Clairvaux, Éloge de la nouvelle chevalerie, Liber ad milites Templi De laude novae militiae, P. A. O à Sources Chrétiennes,col. 367, 1990.
- Bernard de Clairvaux, Précepte et la Dispense, Liber de praecepto et dispensatione, P. A. O à Sources Chrétiennes, col. 457, 2000.
- Bernard de Clairvaux, Grâce et le Libre Arbitre, Liber de gratia et de libero arbitrio, P. A. O à Sources Chrétiennes, col. 393, 1993.
- Bernard de Clairvaux, Degrés de l'humilité et de l'orgueil, Liber de gradibus humilitatis et superbia, SBO III, 13-59, 1957-1977.
- Bernard de Clairvaux, Vie de saint Malachie, Vita sancti Malachiae episcopi, P. A. O à Sources Chrétiennes, col. 367, 1990.
- Bernard de Clairvaux, Sermon lors de la mort de Malachie, Sermo in transitu sancti Malachiae, P. A. O à Sources Chrétiennes, col. 367, 1990
- Bernard de Clairvaux,Sentences, Sententiae, SBO VI-2, 1-256, 1957-1977.
- Bernard de Clairvaux, Conversion (Aux clercs sur la conversion), Sermo de conuersione ad clericos, P. A. O à Sources Chrétiennes, col. 457, 2000.
- Bernard de Clairvaux, Louange de la Vierge Mère, Homiliae super Missus est (In laudibus Virginis Matris), P. A. O à Sources Chrétiennes, col. 390, 1993.
- Bernard de Clairvaux, Sermon pour la Nativité de la Bienheureuse Vierge Marie, Sermo in natiuitate Beatae Mariae Virginis, SBO V, 275-288, 1957-1977.
- Bernard de Clairvaux, Sermons pour la naissance de saint Victor, Sermones varii : In natali sancti Victoris, col. 526, 2010.

BIBLICAL REUSE EXAMPLES

more literal	Bible verse	Bernard reuse
Proverbs 18 3	**impius cum in profundum venerit** peccatorum **contemnit** sed sequitur eum ignominia et obprobrium (*When the wicked man is come into the depth of sins, also contempt comes but ignominy and reproach follow him*)	**Impius** , **cum venerit in profundum** malorum , **contemnit** (*When the wicked man is come into the depth of evil*)

less literal	Bible verse	Clement reuse
1Cor 13 13	νυνὶ δὲ μένει **πίστις** , **ἐλπίς** , **ἀγάπη** , **τὰ τρία ταῦτα** μείζων δὲ τούτων ἡ ἀγάπη (*And now remain faith, hope, love, these three; but the greatest of those is love.*)	**πίστει** καὶ **ἐλπίδι** καὶ **ἀγάπη** (*faith, and hope, and love - in dative case*)
		ἀγάπην , **πίστιν** , **ἐλπίδα** (*love, faith, hope - in accusative case*)
		μένει δὲ τὰ τρία ταῦτα , **πίστις** , **ἐλπίς** , **ἀγάπη** · μείζων δὲ ἐν τούτοις ἡ ἀγάπη (*and remain these three, faith, hope, love; but the greatest among them is love*)

non-literal	Bible verse	Clement reuse
Mt 12 35	ὁ ἀγαθὸς ἄνθρωπος ἐκ τοῦ ἀγαθοῦ θησαυροῦ ἐκβάλλει ἀγαθά , καὶ ὁ **πονηρὸς** ἄνθρωπος ἐκ τοῦ πονηροῦ θησαυροῦ ἐκβάλλει **πονηρά** . (*A good man out of good storage brings out good things , and an evil man out of the evil storage brings evil things .*)	Ψυχῆς , τὰ δὲ ἐκτός , κἂν μὲν ἡ ψυχὴ χρῆται καλῶς , καλὰ καὶ ταῦτα δοκεῖ , ἐὰν δὲ **πονηρῶς** , **πονηρά** , ὁ κελεύων ἀπαλλοτριοῦν τὰ ὑπάρχοντα (*[are whitin the] soul, and some are out, and if the soul uses them good, those things are also thought of as good, but if [they are used as] bad, [they are thought of as] bad; he who commands the renouncement of possessions*)

528 50 2 Philippiens 1 23

Bible: coartor autem e duobus desiderium habens dissolvi et cum Christo esse multo magis melius

(I have the desire to depart and be with Christ is far better)

Reuse: dissolvi desiderant *(I wished to be dissolved)*

classification: close-verbatim

The theme this presentation is based on is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. Changes to the theme are the work of eTRAP.