There are several impressive taxonomies that lists types of LR processes (Ullman, 1957), but these are mainly descriptive and make little attempt to predict what kind of lexical change might happen for a given concept. Some proposed general hypotheses are that nouns are replaced more readily than verbs, that more frequent words are insulated from replacement (Pagel et al., 2007) and that rich synonym networks speed up replacement (Vejdemo and Hörberg, 2016), hypotheses that we can investigate using quantitative approaches on large scale, diachronic text upon completion of this project.

Previous work on automatic detection of LR has been very limited and mainly focused on named entity changes. The interest has mostly been from an information retrieval (IR) perspective (Anand et al. (2012); Berberich et al. (2007, 2009); Morsy and Karypis (2016)). Kanhabua and Nørvåg (2010) find semantically related named entities using Wikipedia links, limiting the method to modern, common domain entities and evaluate indirectly in an IR setting. Kaluarachchi et al. (2010) propose to find named entities via linked verbs that relate over time, referring the problem to diachronic linking of verbs. These attempts are computationally expensive, as they require recurrent computations because the target time is unknown beforehand. In our previous work, Tahmasebi et al. (2012), we rely on bursts in frequency to detect time periods in which we search for name change thus eliminating recurrent computation. In all work, changes are only found pairwise, senses are not differentiated and there is no validity period associated with each name. By finding word replacements after having found word sense changes, we can generalize beyond names and overcome many of these obstacles.

The plan

Automatic detection of lexical replacements over time has only previously targeted named entities. Previous work like Zhang et al. (2016) and Berberich et al. (2009) have used word context rather than sense information, which is suitable for (unambiguous) named entities but not for words in general. Contrary to word sense change, we are targeting senses that are stable over time, so we require the induced senses to exhibit only minor variation to represent the same underlying sense. The problem requires (i) deriving word senses; (ii) tracking word senses over time; and (iii) linking words to word senses to find a word that has been used to replace another for a given sense (not necessarily for all senses of a word).

We will investigate multi-sense embeddings for labeling, extend current word pairs to create chains of change, e.g. $\textit{ipod} \xleftarrow[2012-2001]{} \textit{mp3 player} \xleftarrow[2001-1996]{} \textit{minidisc} \xleftarrow[1996-1992]{} \textit{discman} \xleftarrow[1984-1979]{} \textit{walkman}$ for the word sense of *mobile music player*; and assign validity periods for follow up applications (e.g., Information retrieval) and understanding. We will analyze systematic errors to automatically reduce false positives, e.g., *ear phones, tape, disc, Sony, Apple* for the above, thus rendering the results useless.

Non-automatic research is producing interesting hypotheses on word replacement, identifying replacement affecting factors such as frequency (Pagel et al., 2007) and synonym network density (Vejdemo and Hörberg, 2016) but much of this is based on small over-used databases of core vocabulary. We will address these hypotheses with large, diachronic datasets. For research on language change in general, we will investigate methods to link SC and LR changes (as well as spelling variations) to present all changes to a word and other words in its semantic field, in a scrollable and clickable map with links to relevant text passages, with the aim that it should be understandable and searchable.

A. Anand, S. Bedathur, K. Berberich, and R. Schenkel. Index maintenance for time-travel text search. In *SIGIR*, 2012.

K. Berberich, S. Bedathur, T. Neumann, and G. Weikum. A time machine for text search. In *SIGIR*, 2007.

K. Berberich, S. J. Bedathur, M. Sozio, and G. Weikum. Bridging the Terminology Gap in Web Archive Search. In *WebDB'09 workshop*, 2009.

A. C. Kaluarachchi, A. S. Varde, S. J. Bedathur, G. Weikum, J. Peng, and A. Feldman. Incorporating terminology evolution for query translation in text retrieval with association rules. In *CIKM*, 2010.

N. Kanhabua and K. Nørvåg. Exploiting time-based synonyms in searching document archives. In *JCDL*, 2010.

S. Morsy and G. Karypis. Accounting for language changes over time in document similarity search. *Trans. on Information Systems*, 35(1):1, 2016.

M. Pagel, Q. D. Atkinson, and A. Meade. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449:717–720, 2007.

N. Tahmasebi, G. Gossen, N. Kanhabua, H. Holzmann, and T. Risse. NEER: An Unsupervised Method for Named Entity Evolution Recognition. In *COLING*, 2012.

S. Ullman. *The principles of semantics.* Blackwell Publishers, Oxford, 1957.

S. Vejdemo and T. Hörberg. Semantic Factors Predict the Rate of Lexical Replacement of Content Words. *PLOS ONE*, pages 1–15, Jan. 2016.

Y. Zhang, A. Jatowt, S. S. Bhowmick, and K. Tanaka. The past is not a foreign country: Detecting semantically similar terms across time. *KDE*, 2016.