

That’s sick dude!:

Automatic identification of word sense change across different timescales

Sunny Mitra¹, Ritwik Mitra¹, Martin Riedl²,
Chris Biemann², Animesh Mukherjee¹, Pawan Goyal¹

¹Dept. of Computer Science and Engineering,
Indian Institute of Technology Kharagpur, India – 721302

² FG Language Technology, Computer Science Department, TU Darmstadt, Germany

¹{sunnym, ritwikm, animeshm, pawang}@cse.iitkgp.ernet.in

²{riedl, biem}@cs.tu-darmstadt.de

Abstract

In this paper, we propose an unsupervised method to identify noun sense changes based on rigorous analysis of time-varying text data available in the form of millions of digitized books. We construct distributional thesauri based networks from data at different time points and cluster each of them separately to obtain word-centric sense clusters corresponding to the different time points. Subsequently, we compare these sense clusters of two different time points to find if (i) there is birth of a new sense or (ii) if an older sense has got split into more than one sense or (iii) if a newer sense has been formed from the joining of older senses or (iv) if a particular sense has died. We conduct a thorough evaluation of the proposed methodology both manually as well as through comparison with WordNet. Manual evaluation indicates that the algorithm could correctly identify 60.4% birth cases from a set of 48 randomly picked samples and 57% split/join cases from a set of 21 randomly picked samples. Remarkably, in 44% cases the birth of a novel sense is attested by WordNet, while in 46% cases and 43% cases split and join are respectively confirmed by WordNet. Our approach can be applied for lexicography, as well as for applications like word sense disambiguation or semantic search.

1 Introduction

Two of the fundamental components of a natural language communication are word sense discovery (Jones, 1986) and word sense disambiguation (Ide and Veronis, 1998). While discovery corresponds to acquisition of vocabulary, disambiguation forms the basis of understanding. These

two aspects are not only important from the perspective of developing computer applications for natural languages but also form the key components of language evolution and change.

Words take different senses in different contexts while appearing with other words. Context plays a vital role in disambiguation of word senses as well as in the interpretation of the actual meaning of words. For instance, the word “bank” has several distinct interpretations, including that of a “financial institution” and the “shore of a river.” Automatic discovery and disambiguation of word senses from a given text is an important and challenging problem which has been extensively studied in the literature (Jones, 1986; Ide and Veronis, 1998; Schütze, 1998; Navigli, 2009). However, another equally important aspect that has not been so far well investigated corresponds to one or more changes that a word might undergo in its sense. This particular aspect is getting increasingly attainable as more and more time-varying text data become available in the form of millions of digitized books (Goldberg and Orwant, 2013) gathered over the last centuries. As a motivating example one could consider the word “sick” – while according to the standard English dictionaries the word is normally used to refer to some sort of illness, a new meaning of “sick” referring to something that is “crazy” or “cool” is currently getting popular in the English vernacular. This change is further interesting because while traditionally “sick” has been associated to something negative in general, the current meaning associates positivity with it. In fact, a rock band by the name of “Sick Puppies” has been founded which probably is inspired by the newer sense of the word sick. The title of this paper has been motivated by the above observation. Note that this phenomena of change in word senses has existed ever since the beginning of human communication (Bamman and Crane, 2011; Michel et

al., 2011; Wijaya and Yeniterzi, 2011; Mihalcea and Nastase, 2012); however, with the advent of modern technology and the availability of huge volumes of time-varying data it now has become possible to automatically track such changes and, thereby, help the lexicographers in word sense discovery, and design engineers in enhancing various NLP/IR applications (e.g., disambiguation, semantic search etc.) that are naturally sensitive to change in word senses.

The above motivation forms the basis of the central objective set in this paper, which is to devise a completely unsupervised approach to track noun sense changes in large texts available over multiple timescales. Toward this objective we make the following contributions: (a) devise a time-varying graph clustering based sense induction algorithm, (b) use the time-varying sense clusters to develop a split-join based approach for identifying new senses of a word, and (c) evaluate the performance of the algorithms on various datasets using different suitable approaches along with a detailed error analysis. Remarkably, comparison with the English WordNet indicates that in 44% cases, as identified by our algorithm, there has been a birth of a completely novel sense, in 46% cases a new sense has split off from an older sense and in 43% cases two or more older senses have merged in to form a new sense.

The remainder of the paper is organized as follows. In the next section we present a short review of the literature. In Section 3 we briefly describe the datasets and outline the process of co-occurrence graph construction. In Section 4 we present an approach based on graph clustering to identify the time-varying sense clusters and in Section 5 we present the split-merge based approach for tracking word sense changes. Evaluation methods are summarized in Section 6. Finally, conclusions and further research directions are outlined in Section 7.

2 Related work

Word sense disambiguation as well as word sense discovery have both remained key areas of research right from the very early initiatives in natural language processing research. Ide and Veronis (1998) present a very concise survey of the history of ideas used in word sense disambiguation; for a recent survey of the state-of-the-art one can refer to (Navigli, 2009). Some of the first attempts

to automatic word sense discovery were made by Karen Spärck Jones (1986); later in lexicography, it has been extensively used as a pre-processing step for preparing mono- and multi-lingual dictionaries (Kilgarriff and Tugwell, 2001; Kilgarriff, 2004). However, as we have already pointed out that none of these works consider the temporal aspect of the problem.

In contrast, the current study, is inspired by works on language dynamics and opinion spreading (Mukherjee et al., 2011; Maity et al., 2012; Loreto et al., 2012) and automatic topic detection and tracking (Allan et al., 1998). However, our work differs significantly from those proposed in the above studies. Opinion formation deals with the self-organisation and emergence of shared vocabularies whereas our work focuses on how the different senses of these vocabulary words change over time and thus become “out-of-vocabulary”. Topic detection involves detecting the occurrence of a new event such as a plane crash, a murder, a jury trial result, or a political scandal in a stream of news stories from multiple sources and tracking is the process of monitoring a stream of news stories to find those that track (or discuss) the same event. This is done on shorter timescales (hours, days), whereas our study focuses on larger timescales (decades, centuries) and we are interested in common nouns, verbs and adjectives as opposed to events that are characterized mostly by named entities. Other similar works on dynamic topic modelling can be found in (Blei and Lafferty, 2006; Wang and McCallum, 2006). Google books n-gram viewer¹ is a phrase-usage graphing tool which charts the yearly count of selected letter combinations, words, or phrases as found in over 5.2 million digitized books. It only reports frequency of word usage over the years, but does not give any correlation among them as e.g., in (Heyer et al., 2009), and does not analyze their senses.

A few approaches suggested by (Bond et al., 2009; Pääkkö and Lindén, 2012) attempt to augment WordNet synsets primarily using methods of annotation. Another recent work by Cook et al. (2013) attempts to induce word senses and then identify novel senses by comparing two different corpora: the “focus corpora” (i.e., a recent version of the corpora) and the “reference corpora” (older version of the corpora). However, this method is limited as it only considers two time points to

¹<https://books.google.com/ngrams>

identify sense changes as opposed to our approach which is over a much larger timescale, thereby, effectively allowing us to track the points of change and the underlying causes. One of the closest work to what we present here has been put forward by (Tahmasebi et al., 2011), where the authors analyze a newspaper corpus containing articles between 1785 and 1985. The authors mainly report the frequency patterns of certain words that they found to be candidates for change; however a detailed cause analysis as to why and how a particular word underwent a sense change has not been demonstrated. Further, systematic evaluation of the results obtained by the authors has not been provided.

All the above points together motivated us to undertake the current work where we introduce, for the first time, a completely unsupervised and automatic method to identify the change of a word sense and the cause for the same. Further, we also present an extensive evaluation of the proposed algorithm in order to test its overall accuracy and performance.

3 Datasets and graph construction

In this section, we outline a brief description of the dataset used for our experiments and the graph construction procedure. The primary source of data have been the millions of digitized books made available through the Google Book project (Goldberg and Orwant, 2013). The Google Book syntactic n-grams dataset provides dependency fragment counts by the years. However, instead of using the plain syntactic n-grams, we use a far richer representation of the data in the form of a distributional thesaurus (Lin, 1997; Rychlý and Kilgarriff, 2007). In specific, we prepare a distributional thesaurus (DT) for each of the time periods separately and subsequently construct the required networks. We briefly outline the procedure of thesauri construction here referring the reader to (Riedl and Biemann, 2013) for further details. In this approach, we first extract each word and a set of its context features, which are formed by labeled and directed dependency parse edges as provided in the dataset. Following this, we compute the frequencies of the word, the context and the words along with their context. Next we calculate the lexicographer’s mutual information LMI (Kilgarriff, 2004) between a word and its features and retain only the top 1000 ranked features for ev-

ery word. Finally, we construct the DT network as follows: each word is a node in the network and the edge weight between two nodes is defined as the number of features that the two corresponding words share in common.

4 Tracking sense changes

The basic idea of our algorithm for tracking sense changes is as follows. If a word undergoes a sense change, this can be detected by comparing its senses obtained from two different time periods. Since we aim to detect this change automatically, we require distributional representations corresponding to word senses for different time periods. We, therefore, utilize the basic hypothesis of unsupervised sense induction to induce the sense clusters over various time periods and then compare these clusters to detect sense change. The basic premises of the ‘unsupervised sense induction’ are briefly described below.

4.1 Unsupervised sense induction

We use the co-occurrence based graph clustering framework introduced in (Biemann, 2006). The algorithm proceeds in three basic steps. Firstly, a co-occurrence graph is created for every target word found in DT. Next, the neighbourhood/ego graph is clustered using the Chinese Whispers (CW) algorithm (see (McAuley and Leskovec, 2012) for similar approaches). The algorithm, in particular, produces a set of clusters for each target word by decomposing its open neighborhood. We hypothesize that each different cluster corresponds to a particular sense of the target word. For a detailed description, the reader is referred to (Biemann, 2011).

If a word undergoes sense change, this can be detected by comparing the sense clusters obtained from two different time periods by the algorithm outlined above. For this purpose, we use statistics from the DT corresponding to two different time intervals, say tv_i and tv_j . We then run the sense induction algorithm over these two different datasets. Now, for a given word w that appears in both the datasets, we get two different set of clusters, say C_i and C_j . Without loss of generality, let us assume that our algorithm detects m sense clusters for the word w in tv_i and n sense clusters in tv_j . Let $C_i = \{s_{i1}, s_{i2}, \dots, s_{im}\}$ and $C_j = \{s_{j1}, s_{j2}, \dots, s_{jn}\}$, where s_{kz} denotes z^{th} sense cluster for word w during time interval tv_k .

We next describe our algorithm for detecting sense change from these sets of sense clusters.

4.2 Split, join, birth and death

We hypothesize that word w can undergo sense change from one time interval (tv_i) to another (tv_j) as per one of the following scenarios:

Split A sense cluster s_{iz} in tv_i splits into two (or more) sense clusters, s_{jp_1} and s_{jp_2} in tv_j

Join Two sense clusters s_{iz_1} and s_{iz_2} in tv_i join to make a single cluster s_{jp} in tv_j

Birth A new sense cluster s_{jp} appears in tv_j , which was absent in tv_i

Death A sense cluster s_{iz} in tv_i dies out and does not appear in tv_j

To detect split, join, birth or death, we build an $(m+1) \times (n+1)$ matrix I to capture the intersection between sense clusters of two different time periods. The first m rows and n columns correspond to the sense clusters in tv_i and tv_j respectively. We append an additional row and column to capture the fraction of words, which did not show up in any of the sense clusters in another time interval. So, an element I_{kl} of the matrix

- $1 \leq k \leq m, 1 \leq l \leq n$: denotes the fraction of words in a newer sense cluster s_{jl} , that were also present in an older sense cluster s_{ik} .
- $k = m + 1, 1 \leq l \leq n$: denotes the fraction of words in the sense cluster s_{jl} , that were not present in any of the m clusters in tv_i .
- $1 \leq k \leq m, l = n + 1$: denotes the fraction of words in the sense cluster s_{ik} , that did not show up in any of the n clusters in tv_j .

Thus, the matrix I captures all the four possible scenarios for sense change. Since we can not expect a perfect split, birth etc., we used certain threshold values to detect if a candidate word is undergoing sense change via one of these four cases. In Figure 1, as an example, we illustrate the birth of a new sense for the word ‘compiler’.

4.3 Multi-stage filtering

To make sure that the candidate words obtained via our algorithm are meaningful, we applied multi-stage filtering to prune the candidate word

list. The following criterion were used for the filtering:

Stage 1 We utilize the fact that the CW algorithm is non-deterministic in nature. We apply CW three times over the source and target time intervals. We obtain the candidate word lists using our algorithm for the three runs, then take the intersection to output those words, which came up in all the three runs.

Stage 2 From the above list, we retain only those candidate words, which have a part-of-speech tag ‘NN’ or ‘NNS’, as we focus on nouns for this work.

Stage 3 We sort the candidate list obtained in Stage 2 as per their occurrence in the first time period. Then, we remove the top 20% and the bottom 20% words from this list. Therefore, we consider the *torso* of the frequency distribution which is the most informative part for this type of an analysis.

5 Experimental framework

For our experiments, we utilized DTs created for 8 different time periods: 1520-1908, 1909-1953, 1954-1972, 1973-1986, 1987-1995, 1996-2001, 2002-2005 and 2006-2008 (Riedl et al., 2014). The time periods were set such that the amount of data in each time period is roughly the same. We will also use T_1 to T_8 to denote these time periods. The parameters for CW clustering were set as follows. The size of the neighbourhood (N) to be clustered was set to 200. The parameter n regulating the edge density in this neighbourhood was set to 200 as well. The parameter a was set to lin , which corresponds to favouring smaller clusters by hub downweighting². The threshold values used to detect the sense changes were as follows. For birth, at least 80% words of the target cluster should be novel. For split, each split cluster should have at least 30% words of the source cluster and the total intersection of all the split clusters should be $> 80\%$. The same parameters were used for the join and death case with the interchange of source and target clusters.

5.1 Signals of sense change

Making comparisons between all the pairs of time periods gave us 28 candidate words lists. For

²data available at http://sf.net/p/jobimtext/wiki/LREC2014_Google_DT/

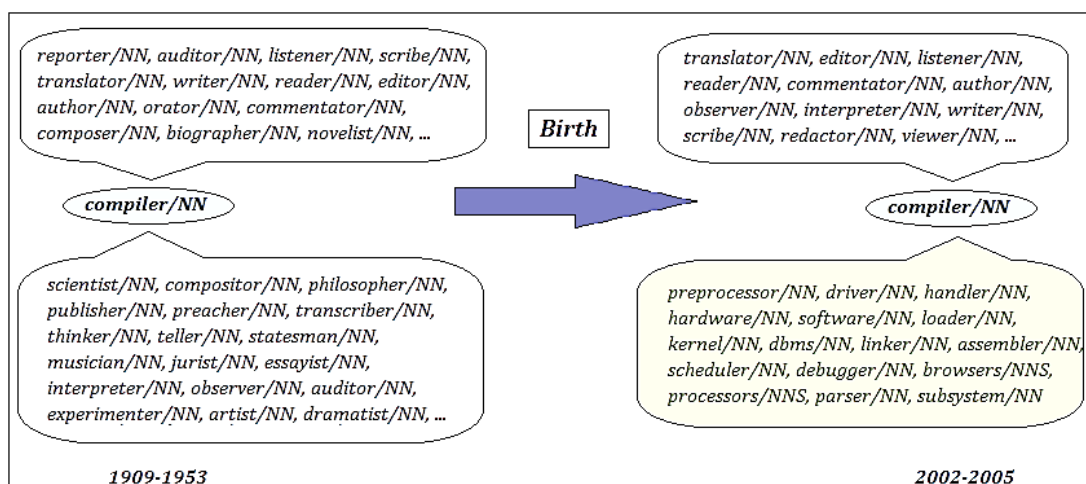


Figure 1: Example of the birth of a new sense for the word ‘compiler’

each of these comparison, we applied the multi-stage filtering to obtain the pruned list of candidate words. Table 1 provides some statistics about the number of candidate words obtained corresponding to the birth case. The rows correspond to the source time-period and the columns correspond to the target time periods. An element of the table shows the number of candidate words obtained by comparing the corresponding source and target time periods.

Table 1: Number of candidate birth senses between all time periods

	T_2	T_3	T_4	T_5	T_6	T_7	T_8
T_1	2498	3319	3901	4220	4238	4092	3578
T_2		1451	2330	2789	2834	2789	2468
T_3			917	1460	1660	1827	1815
T_4				517	769	1099	1416
T_5					401	818	1243
T_6						682	1107
T_7							609

The table clearly shows a trend. For most of the cases, the number of candidate birth senses tends to increase as we go from left to right. Similarly, this number decreases as we go down in the table. This is quite intuitive since going from left to right corresponds to increasing the gap between two time periods while going down corresponds to decreasing this gap. As the gap increases (decreases), one would expect more (less) new senses coming in. Even while moving diagonally, the candidate words tend to decrease as we move downwards. This corresponds to the fact that the number of years in the time periods de-

creases as we move downwards, and therefore, the gap also decreases.

5.2 Stability analysis & sense change location

Formally, we consider a sense change from tv_i to tv_j **stable** if it was also detected while comparing tv_i with the following time periods tv_k s. This number of subsequent time periods, where the same sense change is detected, helps us to determine the **age** of a new sense. Similarly, for a candidate sense change from tv_i to tv_j , we say that the **location** of the sense change is tv_j if and only if that sense change does not get detected by comparing tv_i with any time interval tv_k , intermediate between tv_i and tv_j .

Table 1 gives a lot of candidate words for sense change. However, not all the candidate words were stable. Thus, it was important to prune these results using stability analysis. Also, it is to be noted that these results do not pin-point to the exact time-period, when the sense change might have taken place. For instance, among the 4238 candidate birth sense detected by comparing T_1 and T_6 , many of these new senses might have come up in between T_2 to T_5 as well. We prune these lists further based on the stability of the sense, as well as to locate the approximate time interval, in which the sense change might have occurred.

Table 2 shows the number of stable (at least twice) senses as well as the number of stable sense changes located in that particular time period. While this decreases recall, we found this to be beneficial for the accuracy of the method.

Once we were able to locate the senses as well as to find the age of the senses, we attempted to

Table 2: Number of candidate birth senses obtained for different time periods

	T_2	T_3	T_4	T_5	T_6	T_7
T_1	2498	3319	3901	4220	4238	4092
stable	537	989	1368	1627	1540	1299
located	537	754	772	686	420	300
T_2		1451	2330	2789	2834	2789
stable		343	718	938	963	810
located		343	561	517	357	227

select some representative words and plotted them on a timeline as per the birth period and their age in Figure 2. The source time period here is 1909-1953.

6 Evaluation framework

During evaluation, we considered the clusters obtained using the 1909-1953 time-slice as our reference and attempted to track sense change by comparing these with the clusters obtained for 2002-2005. The sense change detected was categorized as to whether it was a new sense (birth), a single sense got split into two or more senses (split) or two or more senses got merged (join) or a particular sense died (death). We present a few instances of the resulting clusters in the paper and refer the reader to the supplementary material³ for the rest of the results.

6.1 Manual evaluation

The algorithm detected a lot of candidate words for the cases of birth, split/join as well as death. Since it was difficult to go through all the candidate sense changes for all the comparisons manually, we decided to randomly select some candidate words, which were flagged by our algorithm as undergoing sense change, while comparing 1909-1953 and 2002-2005 DT. We selected 48 random samples of candidate words for birth cases and 21 random samples for split/join cases. One of the authors annotated each of the birth cases identifying whether or not the algorithm signalled a true sense change while another author did the same task for the split/join cases. The accuracy as per manual evaluation was found to be 60.4% for the birth cases and 57% for the split/join cases.

Table 3 shows the evaluation results for a few candidate words, flagged due to birth. Columns

³http://cse.iitkgp.ac.in/resgrp/cnerg/acl2014_wordsense/

correspond to the candidate words, words obtained in the cluster of each candidate word (we will use the term ‘birth cluster’ for these words, henceforth), which indicated a new sense, the results of manual evaluation as well as the possible sense this birth cluster denotes.

Table 4 shows the corresponding evaluation results for a few candidate words, flagged due to split or join.

A further analysis of the words marked due to birth in the random samples indicates that there are 22 technology-related words, 2 slangs, 3 economics related words and 2 general words. For the split-join case we found that there are 3 technology-related words while the rest of the words are general. Therefore one of the key observations is that most of the technology related words (where the neighborhood is completely new) could be extracted from our birth results. In contrast, for the split-join instances most of the results are from the general category since the neighborhood did not change much here; it either got split or merged from what it was earlier.

6.2 Automated evaluation with WordNet

In addition to manual evaluation, we also performed automated evaluation for the candidate words. We chose WordNet for automated evaluation because not only does it have a wide coverage of word senses but also it is being maintained and updated regularly to incorporate new senses. We did this evaluation for the candidate birth, join and split sense clusters obtained by comparing 1909-1953 time period with respect to 2002-2005. For our evaluation, we developed an aligner to align the word clusters obtained with WordNet senses. The aligner constructs a WordNet dictionary for the purpose of synset alignment. The CW cluster is then aligned to WordNet synsets by comparing the clusters with WordNet graph and the synset with the maximum alignment score is returned as the output. In summary, the aligner tool takes as input the CW cluster and returns a WordNet synset id that corresponds to the cluster words. The evaluation settings were as follows:

Birth: For a candidate word flagged as birth, we first find out the set of all WordNet synset ids for its CW clusters in the source time period (1909-1953 in this case). Let S_{init} denote the union of these synset ids. We then find WordNet synset id for its birth-cluster, say s_{new} . Then, if $s_{new} \notin$

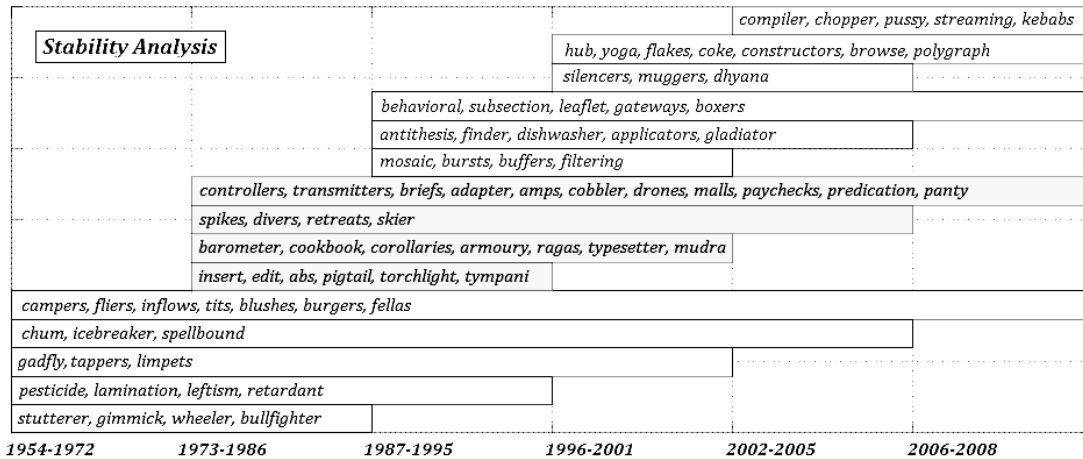


Figure 2: Examples of birth senses placed on a timeline as per their location as well as age

Table 3: Manual evaluation for seven randomly chosen candidate birth clusters between time periods 1909-1953 and 2002-2005

Sl No.	Candidate Word	birth cluster	Evaluation judgement, comments
1	implant	gel, fibre, coatings, cement, materials, metal, filler silicone, composite, titanium, polymer, coating	No, New set of words but similar sense already existed
2	passwords	browsers, server, functionality, clients, workstation printers, software, protocols, hosts, settings, utilities	Yes, New sense related to 'a computer sense'
3	giants	multinationals, conglomerates, manufacturers corporations, competitors, enterprises, companies businesses, brands, firms	Yes, New sense as 'an organization with very great size or force'
4	donation	transplantation, donation, fertilization, transfusions transplant, transplants, insemination, donors, donor ...	Yes, The new usage of donation associated with body organs etc.
5	novice	negro, fellow, emigre, yankee, realist, quaker, teen male, zen, lady, admiring, celebrity, thai, millionaire ...	No, this looks like a false positive
6	partitions	server, printers, workstation, platforms, arrays modules, computers, workstations, kernel ...	Yes, New usage related to the 'computing' domain
7	yankees	athletics, cubs, tigers, sox, bears, braves, pirates cardinals, dodgers, yankees, giants, cardinals ...	Yes, related to the 'New York Yankees' team

S_{init} , it implies that this is a new sense that was not present in the source clusters and we call it a 'success' as per WordNet.

Join: For the join case, we find WordNet synset ids s_1 and s_2 for the clusters obtained in the source time period and s_{new} for the join cluster in the target time period. If $s_1 \neq s_2$ and s_{new} is either s_1 or s_2 , we call it a 'success'.

Split: For the split case, we find WordNet synset id s_{old} for the source cluster and synset ids s_1 and s_2 for the target split clusters. If $s_1 \neq s_2$ and either s_1 , or s_2 retains the id s_{old} , we call it a 'success'.

Table 5 show the results of WordNet based evaluation. In case of birth we observe a success of

Table 5: Results of the automatic evaluation using WordNet

Category	No. of Candidate Words	Success Cases
Birth	810	44%
Split	24	46%
Join	28	43%

44% while for split and join we observe a success of 46% and 43% respectively. We then manually verified some of the words that were deemed as successes, as well as investigated WordNet sense they were mapped to. Table 6 shows some of the words for which the evaluation detected success along with WordNet senses. Clearly, the cluster words correspond to a newer sense for these words

Table 4: Manual evaluation for five randomly chosen candidate split/join clusters between time periods 1909-1953 and 2002-2005

Sl No.	Candidate Word	Source and target clusters
1	intonation (split)	<i>S</i> : whisper, glance, idioms, gesture, chant, sob, inflection, diction, sneer, rhythm, accents ... <i>T</i> ₁ : nod, tone, grimace, finality, gestures, twang, shake, shrug, irony, scowl, twinkle ... <i>T</i> ₂ : accents, phrase, rhythm, style, phonology, diction, utterance, cadence, harmonies ...
		Yes, <i>T</i> ₁ corresponds to intonation in normal conversations while <i>T</i> ₂ corresponds to the use of accents in formal and research literature
2	diagonal (split)	<i>S</i> : coast, edge, shoreline, coastline, border, surface, crease, edges, slope, sides, seaboard ... <i>T</i> ₁ : circumference, center, slant, vertex, grid, clavicle, margin, perimeter, row, boundary .. <i>T</i> ₂ : border, coast, seaboard, seashore, shoreline, waterfront, shore, shores, coastline, coasts
		Yes, the split <i>T</i> ₁ is based on mathematics where as <i>T</i> ₂ is based on geography
3	mantra (join)	<i>S</i> ₁ : sutra, stanza, chanting, chants, commandments, monologue, litany, verse, verses ... <i>S</i> ₂ : praise, imprecation, benediction, praises, curse, salutation, benedictions, eulogy ... <i>T</i> : blessings, spell, curses, spells, rosary, prayers, blessing, prayer, benediction ...
		Yes, the two seemingly distinct senses of mantra - a contextual usage for chanting and prayer (<i>S</i> ₁) and another usage in its effect - salutations, benedictions (<i>S</i> ₂) have now merged in <i>T</i> .
4	continuum (split)	<i>S</i> : circumference, ordinate, abscissa, coasts, axis, path, perimeter, arc, plane axis ... <i>T</i> ₁ : roadsides, corridors, frontier, trajectories, coast, shore, trail, escarpment, highways ... <i>T</i> ₂ : arc, ellipse, meridians, equator, axis, axis, plane, abscissa, ordinate, axis, meridian ...
		Yes, the split <i>S</i> ₁ denotes the usage of 'continuum' with physical objects while the split <i>S</i> ₂ corresponds to its usages in mathematics domain.
5	headmaster (join)	<i>S</i> ₁ : master, overseer, councillor, chancellor, tutors, captain, general, principal ... <i>S</i> ₂ : mentor, confessor, tutor, founder, rector, vicar, graduate, counselor, lawyer ... <i>T</i> : chaplain, commander, surveyor, coordinator, consultant, lecturer, inspector ...
		No, it seems a false positive

and the mapped WordNet synset matches the birth cluster to a very high degree.

6.3 Evaluation with a slang list

Slangs are words and phrases that are regarded as very informal, and are typically restricted to a particular context. New slang words come up every now and then, and this plays an integral part in the phenomena of sense change. We therefore decided to perform an evaluation as to how many slang words were being detected by our candidate birth clusters. We used a list of slangs available from the slangcity website⁴. We collected slangs for the years 2002-2005 and found the intersection with our candidate birth words. Note that the website had a large number of multi-word expressions that we did not consider in our study. Further, some of the words appeared as either erroneous or very transient (not existing more than a few months) entities, which had to be removed from the list. All these removal left us with a very little space for comparison; however, despite this we found 25 slangs from the website that were present in our birth results, e.g. 'bum', 'sissy', 'thug', 'dude' etc.

⁴http://slangcity.com/email_archive/index_2003.htm

6.4 Evaluation of candidate death clusters

Much of our evaluation was focussed on the birth sense clusters, mainly because these are more interesting from a lexicographic perspective. Additionally, the main theme of this work was to detect new senses for a given word. To detect a true death of a sense, persistence analysis was required, that is, to verify if the sense was persisting earlier and vanished after a certain time period. While such an analysis goes beyond the scope of this paper, we selected some interesting candidate "death" senses. Table 7 shows some of these interesting candidate words, their death cluster along with the possible vanished meaning, identified by the authors. While these words are still used in a related sense, the original meaning does not exist in the modern usage.

7 Conclusions

In this paper, we presented a completely unsupervised method to detect word sense changes by analyzing millions of digitized books archived spanning several centuries. In particular, we constructed DT networks over eight different time windows, clustered these networks and compared these clusters to identify the emergence of novel

Table 6: Example of randomly chosen candidate birth clusters mapped to WordNet

Sl No.	Candidate Word	birth cluster	Synset Id, WordNet sense
1	macro	<i>code, query, handler, program, procedure, subroutine module, script</i>	6582403 , a set sequence of steps, part of larger computer program
2	caller	<i>browser, compiler, sender, routers, workstation, cpu host, modem, router, server</i>	4175147 , a computer that provides client stations with access to files
3	searching	<i>coding, processing, learning, computing, scheduling planning, retrieval, routing, networking, navigation</i>	1144355 , programming: setting an order and time for planned events
4	hooker	<i>bitch, whore, stripper, woman slut, prostitute girl, dancer ...</i>	10485440 , a woman who engages in sexual intercourse for money
5	drones	<i>helicopters, fighters, rockets, flights, planes vehicles, bomber, missions, submarines ...</i>	4264914 , a craft capable of traveling in outer space
6	amps	<i>inverters, capacitor, oscillators, switches, mixer transformer, windings, capacitors, circuits ...</i>	2955247 , electrical device characterized by its capacity to store an electric charge
7	compilers	<i>interfaces, algorithms, programming, software modules, libraries, routines, tools, utilities ...</i>	6566077 , written programs pertaining to the operation of a computer system

Table 7: Some representative examples for candidate death sense clusters

Sl No.	Candidate Word	death cluster	Vanished meaning
1	slop	<i>jeans, velveteen, tweed, woollen, rubber, sealskin, wear oilskin, sheepskin, velvet, calico, deerskin, goatskin, cloth ...</i>	clothes and bedding supplied to sailors by the navy
2	blackmail	<i>subsidy, rent, presents, tributes, money, fine, bribes dues, tolls, contributions, contribution, customs, duties ...</i>	Origin: denoting protection money levied by Scottish chiefs
3	repertory	<i>dictionary, study, compendium, bibliography, lore, directory catalogues, science, catalog, annals, digest, literature ...</i>	Origin: denoting an index or catalog: from late Latin repertorium
4	phrasing	<i>contour, outline, construction, handling, grouping, arrangement structure, modelling, selection, form ...</i>	in the sense 'style or manner of expression': via late Latin Greek phrasis

senses. The performance of our method has been evaluated manually as well as by comparison with WordNet and a list of slang words. Through manual evaluation we found that the algorithm could correctly identify 60.4% birth cases from a set of 48 random samples and 57% split/join cases from a set of 21 randomly picked samples. Quite strikingly, we observe that (i) in 44% cases the birth of a novel sense is attested by WordNet, (ii) in 46% cases the split of an older sense is signalled on comparison with WordNet and (iii) in 43% cases the join of two senses is attested by WordNet. These results might have strong lexicographic implications – even if one goes by very moderate estimates almost half of the words would be candidate entries in WordNet if they were not already part of it. This method can be extremely useful in the construction of lexico-semantic networks for low-resource languages, as well as for keeping lexico-semantic resources up to date in general.

Future research directions based on this work are manifold. On one hand, our method can be used by lexicographers in designing new dictionaries where candidate new senses can be semi-automatically detected and included, thus greatly reducing the otherwise required manual effort.

On the other hand, this method can be directly used for various NLP/IR applications like semantic search, automatic word sense discovery as well as disambiguation. For semantic search, taking into account the newer senses of the word can increase the relevance of the query result. Similarly, a disambiguation engine informed with the newer senses of a word can increase the efficiency of disambiguation, and recognize senses uncovered by the inventory that would otherwise have to be wrongly assigned to covered senses. In addition, this method can be also extended to the ‘NNP’ part-of-speech (i.e., named entities) to identify changes in role of a person/place. Furthermore, it would be interesting to apply this method to languages other than English and to try to align new senses of cognates across languages.

Acknowledgements

AM would like to thank DAAD for supporting the faculty exchange programme to TU Darmstadt. PG would like to thank Google India Private Ltd. for extending travel support to attend the conference. MR and CB have been supported by an IBM SUR award and by LOEWE as part of the research center *Digital Humanities*.

References

- J. Allan, R. Papka and V. Lavrenko. 1998. On-line new event detection and tracking. In proceedings of *SIGIR*, 37–45, Melbourne, Australia.
- D. Bamman and G. Crane. 2011. Measuring Historical Word Sense Variation. In proceedings of *JCDL*, 1–10, New York, NY, USA.
- C. Biemann. 2006. Chinese whispers - an efficient graph clustering algorithm and its application to natural language processing problems. In proceedings of *TextGraphs*, 73–80, New York, USA.
- C. Biemann. 2011. *Structure Discovery in Natural Language*. Springer Heidelberg Dordrecht London New York. ISBN 978-3-642-25922-7.
- D. Blei and J. Lafferty. 2006. Dynamic topic models. In proceedings of *ICML*, 113–120, Pittsburgh, Pennsylvania.
- F. Bond, H. Isahara, S. Fujita, K. Uchimoto, T. Kuribayashi and K. Kanzaki. 2009. Enhancing the Japanese WordNet. In proceedings of *workshop on Asian Language Resources*, 1–8, Suntec, Singapore.
- P. Cook, J. H. Lau, M. Rundell, D. McCarthy, T. Baldwin. 2013. A lexicographic appraisal of an automatic approach for detecting new word senses. In proceedings of *eLex*, 49–65, Tallinn, Estonia.
- Y. Goldberg and J. Orwant. 2013. A dataset of syntactic-ngrams over time from a very large corpus of English books. In proceedings of the *Joint Conference on Lexical and Computational Semantics (*SEM)*, 241–247, Atlanta, GA, USA.
- G. Heyer, F. Holz and S. Teresniak. 2009. Change of topics over time – tracking topics by their change of meaning. In proceedings of *KDIR*, Madeira, Portugal.
- N. Ide and J. Veronis. 1998. Introduction to the special issue on word sense disambiguation: The state of the art. *Computational Linguistics*, 24(1):1–40.
- A. Kilgarriff, P. Rychly, P. Smrz, and D. Tugwell. 2004. The sketch engine. In Proceedings of *EU-RALEX*, 105–116, Lorient, France.
- A. Kilgarriff and D. Tugwell. 2001. Word sketch: Extraction and display of significant collocations for lexicography. In proceedings of *COLLOCATION: Computational Extraction, Analysis and Exploitation*, 32–38, Toulouse, France.
- D. Lin. 1997. Using syntactic dependency as local context to resolve word sense ambiguity. In proceedings of *ACL/EACL*, 64–71, Madrid, Spain.
- V. Loreto, A. Mukherjee and F. Tria. 2012. On the origin of the hierarchy of color names. *PNAS*, 109(18), 6819–6824.
- S. K. Maity, T. M. Venkat and A. Mukherjee. 2012. Opinion formation in time-varying social networks: The case of the naming game. *Phys. Rev. E*, 86, 036110.
- J. McAuley and J. Leskovec. 2012. Learning to discover social circles in ego networks. In proceedings of *NIPS*, 548–556, Nevada, USA.
- J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak and E. L. Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182.
- R. Mihalcea and V. Nastase. 2012. Word epoch disambiguation: finding how words change over time. In proceedings of *ACL*, 259–263, Jeju Island, Korea.
- A. Mukherjee, F. Tria, A. Baronchelli, A. Puglisi and V. Loreto. 2011. Aging in language dynamics. *PLoS ONE*, 6(2): e16677.
- R. Navigli. 2009. Word sense disambiguation: a survey. *ACM Computing Surveys*, 41(2):1–69.
- P. Pääkkö and K. Lindén. 2012. Finding a location for a new word in WordNet. In proceedings of the *Global WordNet Conference*, Matsue, Japan.
- M. Riedl and C. Biemann. 2013. Scaling to large³ data: An efficient and effective method to compute distributional thesauri. In proceedings of *EMNLP*, 884–890, Seattle, Washington, USA.
- M. Riedl, R. Steuer and C. Biemann. 2014. Distributed distributional similarities of Google books over the centuries. In proceedings of *LREC*, Reykjavik, Iceland.
- P. Rychlý and A. Kilgarriff. 2007. An efficient algorithm for building a distributional thesaurus (and other sketch engine developments). In proceedings of *ACL, poster and demo sessions*, 41–44, Prague, Czech Republic.
- H. Schütze. 1998. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.
- K. Spärk-Jones. 1986. *Synonymy and Semantic Classification*. Edinburgh University Press. ISBN 0-85224-517-3.
- N. Tahmasebi, T. Risse and S. Dietze. 2011. Towards automatic language evolution tracking: a study on word sense tracking. In proceedings of *EvoDyn*, vol. 784, Bonn, Germany.
- X. Wang and A. McCallum. 2006. Topics over time: a non-Markov continuous-time model of topical trends. In proceedings of *KDD*, 424–433, Philadelphia, PA, USA.
- D. Wijaya and R. Yeniterzi. 2011. Understanding semantic change of words over centuries. In proceedings of the workshop on *Detecting and Exploiting Cultural Diversity on the Social Web*, 35–40, Glasgow, Scotland, UK.