



Title	The UCD-Net System at SemEval-2020 Task 1: Temporal Referencing with Semantic Network Distances
Authors(s)	Nulty, Paul, Lillis, David
Publication date	2020-12
Publication information	Nulty, Paul, and David Lillis. "The UCD-Net System at SemEval-2020 Task 1: Temporal Referencing with Semantic Network Distances." International Committee for Computational Linguistics, December 2020. https://doi.org/10.18653/v1/2020.semeval-1.13 .
Conference details	The 14th International Workshop on Semantic Evaluation (SemEval 2020), Barcelona, Spain (held online due to coronavirus outbreak), 12 December 2020
Publisher	International Committee for Computational Linguistics
Item record/more information	http://hdl.handle.net/10197/25814
Publisher's version (DOI)	10.18653/v1/2020.semeval-1.13

Downloaded 2026-05-01 23:51:43

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

The UCD-Net System at SemEval-2020 Task 1: Temporal Referencing with Semantic Network Distances

Paul Nulty

School of Computer Science
University College Dublin
paul.nulty@ucd.ie

David Lillis

School of Computer Science
University College Dublin
david.lillis@ucd.ie

Abstract

This paper describes the UCD-Net system entered for SemEval 2020 Task 1: *Unsupervised Lexical Semantic Change Detection*. We propose a novel method based on distance between temporally referenced lexical nodes in a semantic network constructed from a combination of the time specific corpora. We argue for the value of semantic networks as objects for transparent exploratory analysis and visualisation of lexical semantic change, and present an application of an interactive semantic network exploration tool for interpreting specific aspects of subtle changes in conceptual meaning.

1 Introduction

SemEval 2020 Task 1: *Unsupervised Lexical Semantic Change Detection* addresses the problem of identifying lexical semantic change using corpus pairs from different time periods in English, Swedish, German, and Latin (Schlechtweg et al., 2020). Measuring and characterising lexical semantic change is of interest to linguists as well as to researchers in the social sciences and humanities who aim to compare shifts in lexical and conceptual meaning across time periods. In common with word sense disambiguation tasks, the evaluation of systems which aim to measure lexical semantic change is complicated by the lack of obvious concrete objectives and natural sources of ground truth labels. SemEval 2020 Task 1 aims to address this by providing a human-labelled ground truth dataset across multiple languages, indicating the loss and/or gain of word senses, and the overall degree to which each word has changed meaning between the two time periods covered by the corpora.

Many previous approaches to this task have been based on constructing vector space representations of target words in each of the two corpora, and measuring the difference between these representations (Tahmasebi et al., 2018). The system described in this paper differs from these approaches by its use of semantic networks to represent the meanings of words in the corpora. A semantic network in this context refers to a graph, with nodes representing words connected by edges that represent association strengths between them. Such networks have often been used in psychology and cognitive science as models for human lexical or conceptual knowledge (Steyvers and Tenenbaum, 2005). The construction of an overall semantic network to represent the structure of meaning across the corpora from both time periods provides a transparent object for exploration and visualisation of the conceptual structure that target words are located in.

The system described adopts an approach inspired by temporal referencing (Dubossarsky et al., 2019), treating the time-specific earlier and later corpora as a single large corpus, constructing a single semantic network from co-occurrences derived from this overall corpus, but replacing target words (and only target words) with time-specific tokens. The change in meaning of the target words can then be measured using node distance metrics in the overall network. The system worked best on English data, ranking joint fifth and eighth of submitted systems for English. In both tasks, the system performed worst on the Latin and German.

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

Averaged across all four languages, the system placed overall 14th out of 26 systems for which a paper was submitted at the task of ranking target words according to their degree of semantic change. In the binary task of predicting whether or not a target word had gained or lost a sense, the average performance of the system underperformed other systems due to incorrectly backing off to predicting change in many cases where the temporally referenced target nodes were not connected in the network — this was usually the result of sparsity in the network (particularly in Latin) rather than actual semantic change.

Overall, visualisation of the constructed semantic networks indicates that qualitatively important changes in target word associations are discovered, but calibrating node distance measures is particularly difficult in the binary task.

This system description paper outlines the methodology for constructing a semantic network from sentences in the dataset, the network distance metric used, and presents an analysis of the strengths and weaknesses of this approach. In addition, we present an interactive web application for exploring the semantic network and visualising the differences between the target words according to their place in the network. This application is available online and may be useful to other researchers in order to explore the different associations of target word senses across the two time periods in a qualitative manner.

2 Background

2.1 Task and Annotation Details

SemEval 2020 Task 1 consists of two closely related subtasks. For subtask 1, systems must provide binary labels indicating whether or not a target word has changed in meaning. A meaning change is deemed to have occurred if the word has gained or lost any senses. For subtask 2, systems must rank all of the target words according to the degree of lexical semantic change that they have undergone between the two corpora (the earlier corpus and the later corpus).

These subtasks are evaluated independently across the four languages involved. The ground truth labels were assigned according to the procedure described in (Schlechtweg et al., 2018). Annotators were presented with pairs of sentences containing a usage of the target word in the earlier and the later corpus, and asked to score the relatedness of the two usages on a four point scale. Changes in the sense frequency distribution are used as ground truth for the ranking task, and the changes in attestation of senses above a threshold are used for the binary task (Schlechtweg and Walde, 2020).

The task is totally unsupervised: neither the ground truth labels nor the class distributions were available prior to system submission, and the number of submissions per system is limited to ten. A different method of annotation was used for Latin, due to the lack of native speakers. Annotators with good knowledge of Latin were recruited for the task, and a set of words known to have changed meaning from pre-Christian to the Christian era were used alongside control words that did not change. The annotators referenced their judgements to a entries in a sense definition dictionary.

2.2 Motivation and related research

Most previous approaches to the general task of lexical semantic change detection begin by constructing individual representations for target words in each of two reference corpora, and then measure the difference between the earlier and later representations. It is common to evaluate the change measure by selecting terms whose representations differ the most and comparing to domain knowledge of concepts in the corpora concerned. This can validate clear cases where new senses have clearly been gained over time, especially where a new sense becomes dominant over the earlier sense.

However, in practice, researchers in humanities and social science are often interested not only in wholesale meaning change resulting from technological advances or dramatic social change, but also in subtle conceptual distinctions in terms that have important meanings in social or political debate. In these cases, the overall degree of change may be relatively small, but differences in the individual semantic relations that drive overall representation change might have important implications. Close reading is the usual method by which these changes are studied, and in political theory the language used to describe these ‘essentially contested concepts’ (Gallie, 1955), which differ in their conceptions

according to ideological and social differences, often suggests a network representation, e.g. “skeletons of rival patterns of thought” (Gray, 1977).

(Freeden, 1994) emphasises that individuals or communities’ particular versions of general concepts (‘conceptions’) can be empirically ascertainable and describable’. This chimes with the descriptive goal and empirical methods commonly now employed by lexicographers to capture the meaning of words in general. Freedden and others (e.g. (Finlayson, 2007; Oppenheim, 1983)) argue that theorists should investigate the structure of political concepts through their actual usage in text, with reference to structures such as a substantive core and optional peripheral components, or the roles filled by concepts as they are expressed in sentences. This approach has much in common with how word meaning is modelled by computational semantics (Jackendoff, 2010; Pustejovsky et al., 1993). This does not assume that analytic treatment of linguistic context can resolve any ‘true’ or ‘correct’ meaning of a contested concept, but simply that usage reflects meaning as held by the community or ideology that produces the text.

In corpus linguistics and lexicography, perhaps the most widely used tool is the Sketch Engine (Kilgarriff et al., 2014), which presents tables summarising the selectional preferences of terms of interest gathered from dependency parsed corpora. Sketch Engine has been widely deployed in lexicography and the study of language learning, but less often for broader questions in social science (Blinder and Allen, 2015). Network representations of concepts are widely studied in cognitive science (Steyvers and Tenenbaum, 2005). (Lopes et al., 2010) describes a web-based interactive tool for exploring networks derived from bioinformatic data. (Van Atteveldt, 2008) describes an RDF-based system for representing and querying semantic network data from Dutch newspaper articles. (De Deyne et al., 2019) is a word association norm dataset with an accompanying online visualisation tool which allows interactive network exploration.

3 System Overview

3.1 Building the semantic network

In this system we create a semantic network in two steps: (i) constructing a weight (or inverse distance) between every pair of words in the vocabulary of the corpus, and (ii) applying thresholds to determine which pairs of words should be connected by an edge in the final network. In this system, the weight of association between two terms is a positive pointwise mutual information (pPMI) score of the co-occurrence frequency of the word pair in the dataset. We adjust this score with context distribution smoothing (Levy et al., 2015) to account for the tendency of PMI measures to overweight infrequent co-occurrences. We use whole sentences as the co-occurrence window, and construct a single co-occurrence matrix covering both the earlier and later corpus of each language. Following a temporal referencing approach, only the target words have co-occurrence counts specific to the earlier or later corpus. For the rest of the vocabulary, the co-occurrence statistics are computed across the earlier and later corpora combined. For each language except Latin, a short list of stopwords was used to exclude function words and very common terms from the vocabulary. No stopword list for Latin was used due to the smaller nature of the corpus, and difficulty in interpreting the density of the resulting networks in this language.

The co-occurrence matrix is converted to a weighted, undirected network by first computing pPMI scores from the raw co-occurrence frequencies, and then building a network with edges connecting every pair of nodes whose pPMI score exceeds a threshold. This results in a single semantic network for each language, containing a node for every word in the vocabulary, with the exception of the target words which each have two separate nodes, `target_c1` and `target_c2`, representing the target word in the earlier corpus and the target word in the later corpus respectively. As an example, the neighbourhood networks of the target words *donkey* and *stab* is shown in figure 1. The pPMI threshold used to generate the networks for the submitted system was chosen by initially using the score of 90th percentile point in the distribution of scores for all co-occurrence pairs, and then manually tweaking this for each language by using the interactive application to find a value that produced networks that were visually informative or reasonable.

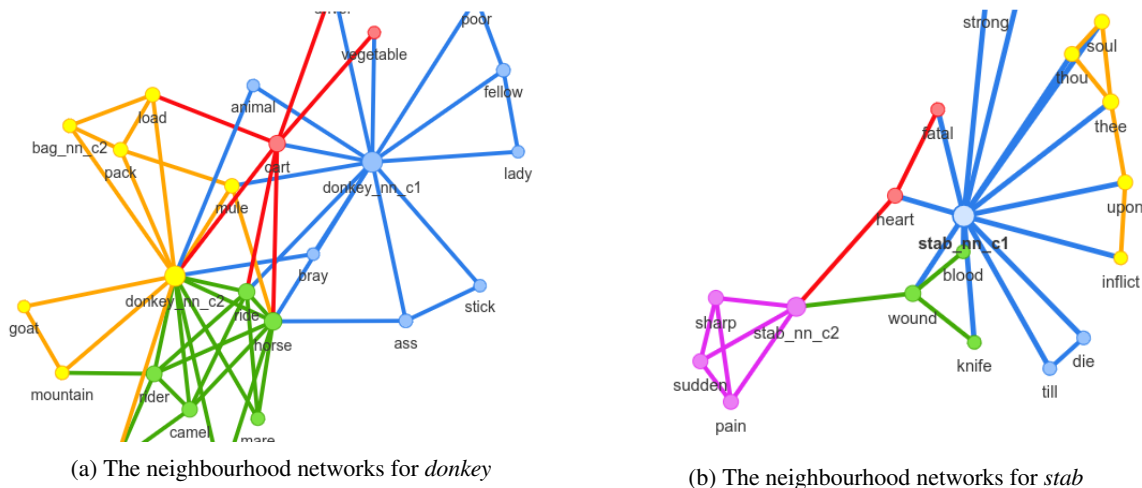


Figure 1: Examples of the neighbourhood networks around temporally referenced nodes from the English data. The temporally reference nodes are indicated with the suffixes *c_1* and *c_2*

3.2 Measuring change between target words

The system aims to measure the degree of semantic change of the target words by measuring the distance between the earlier and later target words in the semantic network. Perhaps the most common method of measuring distance in a weighted network is by finding the shortest path between the two nodes, and summing the distances (or inverse weights) along this path. Exploratory visualisation of the task data suggested that for most of the target words, the earlier and later nodes were within two steps in the network¹, making this distance highly dependent on this single pair of co-occurrence frequencies. Therefore, we chose to use the *resistance distance* metric (Klein and Randić, 1993), as implemented in the python *networkx* package (Vos, 2016), which treats the network as an electrical circuit, and distances as resistors, in order to calculate a distance between the earlier and later nodes that is equivalent to the electrical resistance between the two nodes in such a circuit. The advantage of this metric is that takes into account weights between second order connections in the neighbourhood network of the two nodes of interest.

The system is designed primarily to solve the graded lexical semantic change task, as distinct word senses are not explicitly modelled in this approach. In common with other systems, we submitted scores for the binary classification subtask by predicting that change had occurred in cases where the degree of change exceeded a threshold. This threshold was chosen manually for each language by visually examining the distribution of predicted change scores across the target words (as in the qualitative analysis described in (Schlechtweg et al., 2018)).

3.3 Results

The performance of our system, the best performing system, and the task baselines for both subtasks are shown in Table . The system’s best performance is on the English data, with the average score brought down by a very low correlation in Latin. Exploration of visualisations of the Latin networks suggests that they are much sparser than the other languages, and in many cases even words with low levels of semantic change in the ground truth data were not connected at all in the overall network. This could potentially be improved by calibrating the pPMI thresholds for constructing the network differently.

Our system did not aim to answer the binary classification question directly, and the procedure described in Section 3.1 — visually choosing a cut-off point in the predicted change scores in order to divide change or no change predictions — interacted badly with the distribution of resistance distance scores, as the distribution of the true scores was unknown. As a result, the system performed badly at the subtask 1.

Performance may also have been harmed because no stopword list was used for Latin, with the result

¹the earlier and later target words cannot be connected directly to each other, as they never co-occur in the same sentence in the data

Team	Subtask 1					System	Team	Subtask 2					System
	Avg.	EN	DE	LA	SV			Avg.	EN	DE	LA	SV	
UWB	.687	.622	.750	.700	.677	type	UG_Student_Intern	.527	.422	.725	.412	.547	type
Life-Language	.686	.703	.750	.550	.742	type	Jiaxin & Jinan	.518	.325	.717	.440	.588	type
Jiaxin & Jinan	.665	.649	.729	.700	.581	type	cs2020	.503	.375	.702	.399	.536	type
Count Bas.	.613	.595	.688	.525	.645	-	UCD	.234	.307	.216	.069	.344	graph
Maj. Bas.	.576	.568	.646	.350	.742	-	Count Bas.	.144	.022	.216	.359	-.022	-
UCD	.521	.622	.500	.350	.613	graph	Freq. Bas.	-.083	-.217	.014	.020	-.150	-
Freq. Bas.	.439	.432	.417	.650	.258	-	Maj. Bas.	-	-	-	-	-	-

Table 1: The accuracy (Subtask 1) and Spearman correlation (Subtask 2) corresponding to their best submission in the evaluation phase for our system, the top three systems, and the baselines. Abbreviations: Avg. = average across languages, DE = German, LA = Latin, and SV = Swedish. EN = English.

that the neighbourhood networks around the target terms in the Latin network were dominated by high frequency terms and function words, which influenced the distance metric more than semantically specific links.

4 Visualisation

One of the main motivations of this system is to illustrate an application of interactive semantic network visualisation as a method for exploratory and qualitative analysis of subtle lexical semantic change. The visualisation is implemented as an R Shiny web application. The figures in this section show screenshots from the network representations displayed by the web app. These are neighbourhood or ego graphs of order one, that is, they show nodes within one step of the nodes entered in the search field. The colours are determined by a community detection method that can be selected by the user, but the communities were not used in the change-scoring process.

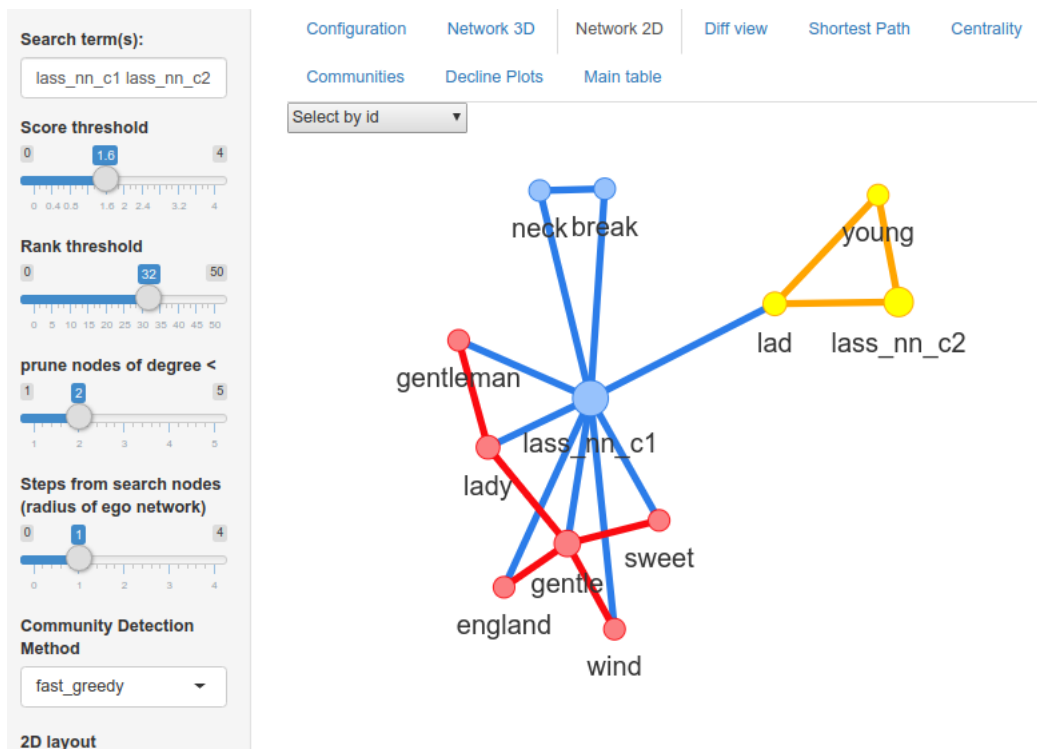


Figure 2: A screenshot of the interactive network exploration interface

In the interactive application (figure 2), the pPMI determining whether an edge exists between two nodes can be changed to increase or decrease the sparsity of the network drawn. A rank threshold can also be applied, whereby edges only exist between the top scoring n words associated with each other word.

The network is drawn using the R `visNetwork` package, using a force-directed algorithm (Fruchterman and Reingold, 1991), which models the network mechanically as repelling particles connected by springs. The result is that in a graph of suitable density and degree, nodes are spaced apart enough to be distinguished, but the edges pull together nodes into clusters that share many relations.

The `visNetwork` package implements a drag, pan, and zoom enabled central widget, and this is combined with input boxes for search terms and sliders for setting the order of the neighbourhood graph, the maximum node degree, and the association score threshold. Network visualisations are often of limited use in static materials — if a network is drawn with enough nodes and density to show important overall patterns, the labelling of the nodes becomes too dense to easily visualise. Interactive visualisation provides a responsive and scalable alternative. The system can be accessed online at: <https://concept-lab.lib.cam.ac.uk/shiny/viewers/viewer-langchange/>, with documentation at <https://concept-lab.lib.cam.ac.uk/viewerdocs/october-docs.html>

5 Conclusion

This system implements a semantic network temporal referencing approach to measuring lexical semantic change from unsupervised data. We present methods for constructing the semantic network, and results from using the resistance distance metric to measure the distance between time-specific target words in the overall semantic network. We also advocate for interactive semantic network exploration as a method to qualitatively analyse subtle distinctions in conceptual meaning that are important for questions usually addressed by close reading, but do not consist in clear changes in word sense distributions.

Acknowledgements

This work was supported by the Enterprise Ireland and European Union Career-FIT programme under the Marie Skodowska-Curie grant agreement No. 713654 ..

References

- Scott Blinder and William L Allen. 2015. Constructing immigrants: Portrayals of migrant groups in british national newspapers, 2010–2012. *International Migration Review*.
- Simon De Deyne, Danielle J Navarro, Amy Perfors, Marc Brysbaert, and Gert Storms. 2019. The small world of words english word association norms for over 12,000 cue words. *Behavior research methods*, 51(3):987–1006.
- Haim Dubossarsky, Simon Hengchen, Nina Tahmasebi, and Dominik Schlechtweg. 2019. Time-out: Temporal referencing for robust modeling of lexical semantic change. *arXiv preprint arXiv:1906.01688*.
- Alan Finlayson. 2007. From beliefs to arguments: Interpretive methodology and rhetorical political analysis. *The British Journal of Politics and International Relations*, 9(4):545–563.
- Michael Freedon. 1994. Political concepts and ideological morphology. *Journal of Political Philosophy*, 2(2):140–164.
- Walter Bryce Gallie. 1955. Essentially contested concepts. In *Proceedings of the Aristotelian society*, volume 56, pages 167–198. JSTOR.
- John N Gray. 1977. On the contestability of social and political concepts. *Political theory*, 5(3):331–348.
- R. Jackendoff. 2010. *Meaning and the lexicon: the parallel architecture, 1975-2010*. Oxford University Press.
- Adam Kilgarriff, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. The sketch engine: ten years on. *Lexicography*, 1(1):7–36.
- Douglas J Klein and Milan Randić. 1993. Resistance distance. *Journal of mathematical chemistry*, 12(1):81–95.
- Omer Levy, Yoav Goldberg, and Ido Dagan. 2015. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3:211–225.
- Christian T Lopes, Max Franz, Farzana Kazi, Sylva L Donaldson, Quaid Morris, and Gary D Bader. 2010. Cytoscape web: an interactive web-based network browser. *Bioinformatics*, 26(18):2347–2348.

- Felix E Oppenheim. 1983. Political concepts: A reconstruction.
- James Pustejovsky, Peter Anick, and Sabine Bergler. 1993. Lexical semantic techniques for corpus analysis. *Comput. Linguist.*, 19:331–358, June.
- Dominik Schlechtweg and Sabine Schulte im Walde. 2020. Simulating lexical semantic change from sense-annotated data. *arXiv preprint arXiv:2001.03216*.
- Dominik Schlechtweg, Sabine Schulte im Walde, and Stefanie Eckmann. 2018. Diachronic usage relatedness (DUREl): A framework for the annotation of lexical semantic change. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 169–174, New Orleans, Louisiana, June. Association for Computational Linguistics.
- Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. 2020. SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. In *To appear in Proceedings of the 14th International Workshop on Semantic Evaluation*, Barcelona, Spain. Association for Computational Linguistics.
- Mark Steyvers and Joshua B Tenenbaum. 2005. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, 29(1):41–78.
- Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2018. Survey of computational approaches to lexical semantic change. *arXiv preprint arXiv:1811.06278*.
- Wouter Van Atteveldt. 2008. Semantic network analysis: Techniques for extracting, representing, and querying media content. *PhD Thesis*.
- VSS Vos. 2016. *Methods for determining the effective resistance*. Ph.D. thesis, Masters thesis, 20 December.