UDK 81’374
81’322.2
Prethodno priopćenje
Rukopis primljen 11. XII. 2019.
Prihvaćen za tisak 18. III. 2020.
doi.org/10.31724/rihjj.46.2.21

Nikola Ljubešić
Jožef Stefan Institute
Jamova cesta 39, SI-1000 Ljubljana
nikola.ljubesic@ijs.si

“DEEP LEXICOGRAPHY” – FAD OR OPPORTUNITY?

In recent years, we are witnessing staggering improvements in various semantic data processing tasks due to the developments in the area of deep learning, ranging from image and video processing to speech processing, and natural language understanding. In this paper, we discuss the opportunities and challenges that these developments pose for the area of electronic lexicography. We primarily focus on the concept of representation learning of the basic elements of language, namely words, and the applicability of these word representations to lexicography. We first discuss well-known approaches to learning static representations of words, the so-called word embeddings, and their usage in lexicography-related tasks such as semantic shift detection, and cross-lingual prediction of lexical features such as concreteness and imageability. We wrap up the paper with the most recent developments in the area of word representation learning in form of learning dynamic, context-aware representations of words, showcasing some dynamic word embedding examples, and discussing improvements on lexicography-relevant tasks of word sense disambiguation and word sense induction.

1. Introduction

In the last few years, we are witnessing great improvements in semantic processing of complex data such as image, video, and audio data. While object detection, one of the fundamental problems in computer vision, has been tackled for decades, the state-of-the-art has improved in recent years dramatically. The top-performing system in 2008 was achieving mean average precision on a series of benchmark task of 21%, with systems in 2014 improving from 33.7% to 59.2%, in 2015 to 73.2%, in 2017 to 76.8%, the current state-of-the-art in 2019 lying at 83.5% (Zou et al. 2019). This represents a 2.5 times improvement on a task in just 5 years. The reason for such incredible improvements is to be found in the area of deep learning, which trains neural networks, consisting of multiple layers (therefore the term "deep"), containing millions, and sometimes even billions of parameters, thereby being able to model complex phenomena such as object detection or speech recognition.

Recently, similar improvements started to emerge on benchmark tests in the area of natural language understanding, the most well-known benchmark being the GLUE (General Language Understanding Evaluation) benchmark[1] (Wang et al. 2018). While it is rather wrong to claim that the level of improvement on benchmark datasets relates to the level of improvement on the problem itself, especially if the task is as complex as natural language understanding, the recently obtained improvements are also quite fascinating. Basic approaches employing static word embeddings, a concept that will be discussed in the first half of this paper, achieve a score of 58.6 on the GLUE benchmark, while the first approaches employing dynamic word embeddings, the concept discussed in the second half of this paper, achieve a score of 66.4. Since the introduction of the benchmark in 2018, improvements were reported on a monthly basis, with the human baseline lying at 87.1, and the current top-performing system achieving a score of 89.7.[2] As with computer vision, these improvements are to be followed back to the area of deep learning. While these are very early days of both benchmarks and models of natural language understanding, and there is a lively discussion going on how well the benchmarks represent the concept of understanding language, which also results in new benchmarks, one of them being the more challenging SuperGLUE (Wang et al. 2019)[3], it is our claim in this paper that the early concepts from the "deep learning revolution" in natural language understanding – static word embeddings – are stable enough to be applied in the downstream of computational lexicography, with dynamic embeddings currently going through significant developments, and their fruits for computational lexicography ripening at a fast pace.

In this paper, we discuss the potential of the deep learning revolution having a positive impact on the area of computational lexicography. More concretely, we

1 https://gluebenchmark.com
2 https://gluebenchmark.com/leaderboard
3 https://super.gluebenchmark.com/leaderboard

explain two basic concepts developed in the previous six years, namely static and dynamic word embeddings and their applicability to the tasks relevant for computational lexicography. While word embeddings are numerical representations of word meaning, the difference between the static and the dynamic embeddings is that (1) the static word embeddings are always the same for a given word, while (2) in the case of dynamic word embeddings, the numerical representation of a word depends on the context in which it occurs.

This paper is structured as follows. In Section 2 we describe the general idea behind representation learning and give a gentle introduction to static word embeddings as a way of representing the meaning of words. In Section 3 we give two examples of using static word embeddings on tasks close to lexicography. In Section 4 we give an even more gentle introduction to dynamic word embeddings and showcase their potential on a few hand-picked usage examples and report the first improvements obtained with this technique on two natural language processing tasks close to lexicography: word sense disambiguation (automatically disambiguating a word in context, e.g., for the context "He robbed a bank recently" correctly predicting that "bank" refers to a financial institution) and word sense induction (automatically inducing various senses of a specific word from, e.g., for the word "bank" identifying at least the "river bank" and the "financial institution" sense distinction).

2. Static word embeddings

One of the first tasks in semantic processing of data such as language or image is to transform the full information present in these data objects, into a representation that is more useful for performing the specific semantic task. Examples of such tasks are identifying objects in an image, transforming speech into text, or translating text from one into another language.

Computers can surely be considered a "tabula rasa" for all human tasks, language understanding included. One of the most obvious data sources from which the language knowledge can be obtained by computers are the large, or rather huge collections of text available nowadays. This fact made room for the rise of the area of distributional modeling of meaning of words (Lenci 2008), which is based on the famous distributional hypothesis popularized by Firth's (1957) quotation: "You shall know a word by the company it keeps". The basic idea behind the distributional modeling of the meaning of words is quite simple: build a model of a word meaning by taking into account all the contexts in which this word occurs in a corpus.

We will give a short example of this approach, taken from Fišer and Ljubešić 2018: let us assume we have a very limited corpus of three sentences, S={"A furry cat runs from a big dog", "The cute dog runs after the red car", "A speeding car hit a cute cat"}, and want to model the words V={"car", "cat", "dog"} based on the words occurring right next to the words we want to model, we can obtain the following co-occurrence matrix:

	big	cute	furry	hit	red	runs	speeding
car	0	0	0	1	1	0	1
cat	0	1	1	0	0	1	0
dog	1	1	0	0	0	1	0

If we consider the sequences of numbers, i.e., rows, for each of our words of interest, V={"car", "cat", "dog"}, to be numerical representations of the meaning of these words, by comparing these representations we see right away that "cat" and "dog" are much more similar (they have five out of seven identical values, the so-called dimensions) than "car" and "cat" (only one identical value, 0 in the "big" dimension) or "car" and "dog" (one identical value as well, 0 in the "furry" dimension). We will not go here into more details of how full models of word embeddings are built, especially not the prediction-based models that are used to produce state-of-the-art static word embeddings but point the interested reader to Fišer and Ljubešić 2018.

The overall takeaway from this section is that static representations of a word are numerical representations based on what words this specific word co-occurs within a corpus. By doing so, we can assume that (1) words of similar meaning will have similar representation and (2) various lexical features of these words will be encoded in these representations, both of which will be shown to be true in the next section.

3. Usage examples for static word embeddings

In this section, we showcase two usages of static word embeddings: one on identifying semantic shift between standard and Internet language in Slovene (Fišer and Ljubešić 2018), and another on predicting the level of concreteness and imageability of words inside, but also across languages (Ljubešić et al. 2018).

3.1. Semantic shift detection

Fišer and Ljubešić (2018) describe an analysis of static word embeddings built on the Gigafida corpus of standard Slovene on one side, and the JANES corpus of computer-mediated communication in Slovene on the other. The analysis is focused on comparing the models of the same word built on one or the other text collection, with the goal of identifying semantic shifts – changes in meaning that occur between these two varieties.

The authors identify three types of major semantic shift: (1) event-based, (2) register-based, and (3) medium-based. An example of the event-based shift is the Slovene word "pirat", unsurprisingly meaning "pirate", which is in standard language primarily used in the negative context of a sea burglar, while in Internet communication the dominant usage is for members of the new political party, mostly in positive contexts. An example of a register-based semantic shift is the Slovene noun "penzion" which in standard Slovene means "guesthouse" but is also used for "retirement" in non-standard language. Finally, an example of a medium-based semantic shift is the noun "sledilec", Slovene for "follower", which goes through a specialization of its meaning in computer-mediated communication, covering almost exclusively only the meaning of a follower on social media.

3.2. Cross-lingual prediction of concreteness and imageability

Ljubešić et al. (2018) investigate the possibility of predicting concreteness (how concrete vs. abstract a word is) and imageability (how easy the concept behind the word can be imagined) of a specific word from the static word embedding of that word.

Regarding the notions of concreteness and imageability, these seem to be close-synonyms, but sometimes they are not. While most abstract concepts are hard to visualize, some call up images, e.g., “torture” calls up an emotional and even visual image. There are concrete things that are hard to visualize too, for example, “abbey” is harder to visualize than “banana” (Tsvetkov et al. 2014).

The authors follow the supervised machine learning paradigm, i.e., they use a lexicon of a few thousand words with known concreteness and imageability and learn to predict these two lexical variables from their static word embeddings. Evaluating the predictor of concreteness and imageability on one English lexicon shows for the predictions to be of high quality with a Spearman correlation coefficient of the true and the predicted estimates of concreteness of 0.872 and imageability of 0.803.

The authors push their approach one important step further by using cross-lingual word embeddings, i.e., word embeddings from different languages, that were transformed in such a way that the known translations of words are as close as possible in the semantic space, expecting for the unknown translations to be very close as well. To give a simple example, if we have the word embedding space separately built for Croatian and English, we can move the Croatian embedding space in such a way that the embeddings of known translations such as “dog” and “pas” are as close as possible. With this transformation of one of the embedding spaces, it is expected that unknown translations will be close in the two spaces, or even further, that the same linguistic information will be encoded in the same part of the embedding space. The concept of cross-lingual static word embeddings is based on the assumption of isomorphism, i.e., that the distances between the same concepts in different languages are the same. Current research shows that this assumption does hold to a specific degree, with typologically more distant language pairs showing less isomorphism (Ormazabal et al. 2019).

The evaluation of cross-lingual concreteness and imageability prediction on the Croatian-English language pair, i.e., training predictors of concreteness and imageability on English words and applying these predictors on Croatian words, using again the Spearman correlation coefficient to quantify the quality of the predictions, shows for both concepts to be highly predictable from static word embeddings even between languages, with concreteness obtaining a score of 0.797 and imageability a score of 0.694.

Some examples of words (written in lowercase) predicted as most and least concrete in different languages are given in Table 1.

Table 1: Examples of words predicted as most and least concrete

German	most abstract: fragwürdig voraussetzen gerechtfertigt erstrebenswert erachten most concrete: schachtel schilder schläger tüte ballen henne beutel teller deckel
Greek	most abstract: υπερβατικός καθοριστικός ουσιαστικός πρακτικός γνωστικός most concrete: γιλέκο δέμα φτυάρι πετσέτα κουτάλι κοχύλι χερούλι πανωφόρι
Finnish	most abstract: käsitteellinen merkityksellinen tarkoituksellinen most concrete: laukko lasinen penkki puntti lastuja vatsastaan tappi seinässä
Russian	most abstract: безусловный безусловное объективный объективность most concrete: рубашка гвозди деревянную шарик кровати молотком

A full lexicon of both concreteness and imageability for 77 languages, i.e. the languages that were present in the cross-lingual static word embedding collection, was published at http://hdl.handle.net/11356/1187.

The take-home message from the two presented studies is the following:

(1) We can easily track the change in the meaning of specific words via static word embeddings by measuring the distance between word embeddings. This distance can be useful for calculating the semantic (dis)similarity between identical (semantic shift) or different (semantic distance) words, in different genres, domains, periods, varieties, etc.

(2) Static word embeddings encode various properties of words and can be used in a supervised machine learning setting to obtain the properties of interest (in our usage example concreteness and imageability) for words for which these properties are not currently available or known.

(3) Cross-lingual embeddings enable the extraction of specific properties not only in one language but also across different languages. Having a lexicon of concreteness in English enabled us to extract a lexicon of concreteness and imageability in 77 different languages.

4. Beyond static embeddings – dynamic word embeddings

Discussing static word embeddings must have given a linguist or lexicographer the chills: we build models of word meaning regardless of the specific contexts in which a word appears, disregarding completely the potential polysemy and homonymy of specific words. Thereby we build one single model for the word "fly", regardless of whether it is a verb or a noun, and whether it represents an animal or a regular part of trousers. While static word embeddings have proven to be highly useful for many tasks, there obviously is room for improvement.

The recent significant improvement came from the idea already heavily used in computer vision: let us build a large model aware of the visual properties of the world, and use this model as a starting point for specific tasks in computer vision, e.g. recognizing wildfires, on top of this large basic model. Such an approach enabled practitioners in computer vision not having the huge resources necessary to build large models to achieve state-of-the results as well, but also can be considered more economical and environment-friendly ("Green AI"; Schwartz et al. 2019) as these basic computer vision models do not have to be built for each task from scratch. Two best-known examples of such pre-trained models for computer vision are Mask-RCNN (He et al. 2017) for object recognition (examples of applications can be inspected here[4]), and OpenPose (Cao et al. 2018) for detecting human body, hand, and facial keypoints (examples of applications can be inspected here[5]).

Recently the same idea has been proposed for natural language understanding. The idea was the following: let us use the vast quantities of text available to train models that understand the regularities of language, and then later tune this model to specific tasks, such as discriminating between word senses in a running text, or inducing various word senses of specific words. What came out of these models was the possibility to extract dynamic word embeddings, i.e., numerical representations of specific words that depend on the context in which the specific word occurred. Here we cannot and do not want to go into details on how these large models are built. For an introduction into the road from static to dynamic word embeddings, we point the interested reader to Smith 2019.

4 https://github.com/matterport/Mask_RCNN
5 https://github.com/CMU-Perceptual-Computing-Lab/openpose

In the remainder of this section, we (1) present a usage example with the BERT model on calculating dynamic word embeddings of words of choice in various contexts and compare the distance of these embeddings and (2) report on the improvements obtained with pre-trained models on the tasks of word sense disambiguation and word sense induction.

4.1. Usage example for comparing dynamic word embeddings

In this subsection, we showcase the BERT model (Devlin et al. 2018) developed by researchers from Google in 2018. In short, the model is a large neural network of up to 24 layers and up to 340 million parameters, trained on the tasks of predicting a masked word in context ("This paper is [MASK] interesting", the probable predictions for the masked token being, among others, "very" and "not") and predicting whether a sentence follows another sentence ("This paper is very interesting. The cows have run from the meadow." being a negative example of sentences following each other, while the sentence pair "This paper is very interesting. I have learned a lot from it." is a positive example).

In the following usage example, we use the BERT model for English and extract dynamic word embeddings for words in a specific context and compare them by calculating the cosine distance between them (the closer the distance to 0, the more similar the representations are, while the closer to 1, the less similar they are), probing thereby whether the contextual specificity is encoded in these embeddings. We start with a sanity check - calculating the distance between the words "car", "cat", and "dog" in a fixed context. As expected, the distance (we calculate between the "cat" and "dog" is 0.094, while the distance between the "car" and the two remaining words is higher, between 0.137 and 0.144.

Moving on to the word "dog", to give a collocational example of meaning change, calculating the distance of the word "dog" in the contexts "The large dog", "The cute dog", and "The hot dog" shows for the cute and large dog to have a distance of 0.087, while the distance of these two living dogs to the similarly called snack of 0.209 and 0.213. We see in this example that the specificity of *dog* in "hot dog", the dog encoding a meal rather than a living dog, has overall been preserved, with almost three times the distance between the living instances of dogs and the meal in comparison to the distance between the two living dogs.

We continue with an example of words being of different parts of speech. In the set of sentences containing the word “work”, “We work”, “Hard work”, “People work” and “Interesting work”, the two most similar “work” instances being “We work” and “People work” (0.094), followed by those of “Hard work” and “Interesting work” (0.113). All other distances lie between 0.172 and 0.196. From this example, we can assess that the difference between the part-of-speech in the word “work” has been preserved in each of the dynamic word embeddings.

This is probably a good time to place caution over such distance-based probes over the dynamic word embeddings. While here we can observe a difference (i.e. distance) regarding the different collocational or part-of-speech usages of specific words in dynamic embeddings, we do not probe whether the difference in the dynamic word embeddings is exactly due to encoding living beings vs. meals, or nouns vs. verbs. Such focused probings over linguistic representations are a highly active area of research (Conneau et al. 2018, Tenney et al. 2019) and are beyond the scope of this paper.

Finally, we give an example of both polysemy and different parts-of-speech with the word “fly”. Calculating the distance of “fly” between the short utterances “Open fly”, “Buzzing fly”, “Fat fly” and “Birds fly”, the closest dynamic representations of the word “fly” are “Buzzing fly” and “Fat fly” (0.074), followed by the distance of “Open fly” and the two previous utterances (0.936 and 0.947), with the largest distance being between these three utterances and the “Birds fly” one, which lie between 0.144 and 0.16.

We believe that through these examples we have shown that (1) dynamic embeddings tend to encode various specificities of a word (part-of-speech, sense, collocation) in a given context and (2) these embeddings should have a positive impact on downstream tasks relevant for computational lexicography, which is what we touch upon in short in the following subsection.

4.2. Improvements obtained via dynamic word embeddings on relevant NLP tasks

On the last shared task focused on word sense disambiguation at SemEval 2015, i.e., the task of classifying each instance of a word into its corresponding sense in a previously known sense inventory, the most-frequent-sense baseline achieved a score (defined between 0 and 100) of 67.1, the best knowledge-based system based on WordNet knowledge achieved a score of 70.3 only (Agirre et al. 2014), while the recently proposed system based on the BERT pre-trained model, the same model we have showcased at the beginning of this section, achieved a score of 80.4 (Huang et al. 2019), removing 33% of the error over the best knowledge-based system on this highly challenging task.

On the last shared task focused on word sense induction at SemEval 2013, i.e., the task of identifying the sense inventory from a list of occurrences of a word, the traditional methods available in 2013 did not manage to surpass the score (defined between 0 adn 100) of 15.92 (Baskaya et al. 2013), while the BERT-based model recently obtained a score of 37.0 (Amrami and Goldberg 2019), which is, again a staggering improvement of more than doubling the previous result.

These developments are just a start of applying the recent advances in natural language processing in the form of pre-trained models for general natural language understanding on tasks relevant for computational lexicography.

The overall take-home messages from this section are the following: (1) the dynamic word embeddings are still in their early phase of development, but (2) they show how to encode the contextual specificities of a word, and (3) therefore, they show strong improvements in lexical-semantic tasks. Therefore, it is our strong belief that each practitioner in the area of computational lexicography should keep their eyes on further developments in the area of pre-trained natural-language-understanding models.

5. Conclusion

In this paper, we introduced two approaches to building representations of the meaning of basic lexical building blocks of language – words. While the first approach builds context-invariant, static representations of words, the second generated representations based on the context in which a word has occurred.

While the first approach – static word embeddings – is much simpler and context-insensitive, we have shown that it is still very useful for identifying semantic shifts and extracting specific lexical features. This approach to calculating meanings of words we consider to be well understood and ready to be applied heavily on various tasks of computational lexicography.

The second approach – dynamic word embeddings calculated from huge language models – shows that it has a huge potential also for the area of computational lexicography, but is still in its infancy. However, given the speed of recent developments in the area of these large language models, the application of this type of approach to problems of computational lexicography can be expected in the near future.

To touch upon the question stated in the title of this paper, "deep lexicography – fad or opportunity", we can conclude that applying results of deep learning on computational lexicography is highly promising and surely is no quickly passing trend. However, as with any new technology, we cannot expect "deep lexicography" to be the magic wand, but just another, much better tool to be used by human lexicographers in describing the lexicon of natural languages. We surely look forward to these developments.

References

Agirre, Eneko; López De Lacalle, Oier; Soroa, Aitor. 2014. Random walks for knowledge-based word sense disambiguation. *Computational Linguistics* 40/1. 57–84. doi.org/10.1162/COLI_a_00164.

Amrami, Asaf; Goldberg, Yoav. 2019. Towards better substitution-based word sense induction. *arXiv* 1905.12598.

Baskaya, Osman; Sert, Enis; Cirik, Volkan; Yuret, Deniz. 2013. Ai-ku: Using substitute vectors and co-occurrence modeling for word sense induction and disambiguation. *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation*. Eds. Manandhar, Suresh; Yuret, Deniz. Association for Computational Linguistics. Atlanta. 300–306.

Cao, Zhe; Hidalgo, Gines; Simon, Tomas; Wei, Shih-En; Sheikh, Yaser. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv* 1812.08008.

Conneau, Alexis; Kruszewski, German; Lample, Guillaume; Barrault, Loïc; Baroni, Marco. 2018. What you can cram into a single vector: Probing sentence embeddings for linguistic properties. *arXiv* 1805.01070.

Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv* 1810.04805.

Firth, John Rupert. 1961. *Papers in Linguistics 1934-1951*. Oxford University Press. London.

Fišer, Darja; Ljubešić, Nikola. 2018. Distributional modelling for semantic shift detection. *International Journal of Lexicography* 32/2. 163–183. doi.org/10.1093/ijl/ecy011.

He, Kaiming; Gkioxari, Georgia; Dollár, Piotr; Girshick, Ross. 2017. Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*. ICCV. Venice. 2961–2969. doi.org/10.1109/ICCV.2017.322.

Huang, Luyao; Sun, Chi; Qiu, Xipeng; Huang, Xuanjing. 2019. GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge. *arXiv* 1908.07245.

Lenci, Alessandro. 2008. Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics* 20/1. 1–31.

Ljubešić, Nikola; Fišer, Darja; Peti-Stantić, Anita. 2018. Predicting Concreteness and Imageability of Words Within and Across Languages via Word Embeddings. *arXiv* 1807.02903.

Ormazabal, Aitor; Artetxe, Mikel; Labaka, Gorka; Soroa, Aitor; Agirre, Eneke. 2019. Analyzing the limitations of cross-lingual word embedding mappings. *arXiv* 1906.05407.

Schwartz, Roy; Dodge, Jesse; Smith, Noah. A.; Etzioni, Oren. 2019. Green AI. *arXiv* 1907.10597.

Smith, Noah A. 2019. Contextual Word Representations: A Contextual Introduction. *arXiv* 1902.06006.

Tenney, Ian et al. 2019. What do you learn from context? probing for sentence structure in contextualized word representations. *arXiv* 1905.06316.

Tsvetkov, Yulia; Boytsov, Leonid; Gershman, Anatole; Nyberg, Eric; Dyer, Chris. 2014. Metaphor detection with cross-lingual model transfer. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Eds. Toutanova, Kristina; Wu, Hua. Association for Computational Linguistics. Baltimore. 248–258. dx.doi.org/10.3115/v1/P14-1024.

Wang, Alex; Singh, Amanpreet; Michael, Julian; Hill, Felix; Levy, Omer; Bowman, Samuel R. 2018. GLUE: A Multi-task Benchmark and Analysis Platform for Natural Language Understanding. *arXiv* 1804.07461. dx.doi.org/10.18653/v1/W18-5446.

Wang, Alex et al. 2019. SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems. *arXiv* 1905.00537.

Zou, Zhengxia; Shi, Zhenwei; Guo, Yuhong; Ye, Jieping. 2019. Object Detection in 20 Years: A Survey. *arXiv* 1905.05055.

„Duboka leksikografija" – pomodnost ili prilika?

Sažetak

Posljednjih smo godina svjedoci velikoga napretka u različitim zadatcima inteligentne obrade podataka koji je posljedica razvoja dubokoga učenja. Ti zadatci uključuju i obradu slike, videa, govora te razumijevanje jezika. U ovome se radu raspravlja o prilikama i izazovima koje taj napredak omogućuje u području digitalne leksikografije.
Veći se dio rada odnosi na učenje prikaza različitih elemenata jezika – riječi, leksema te izjava – i njihovu primjenu u leksikografiji. Prikazuju se dobro poznati postupci učenja statičkih vektorskih prikaza riječi te njihova primjena u zadatcima poput prepoznavanja semantičkih pomaka te predviđanja leksičkih značajka riječi. U radu se dalje govori o višejezičnoj razini učenja prikaza riječi te se rad zaključuje prikazom novijih postignuća u području strojnoga razumijevanja jezika – dinamičkih, kontekstnih prikaza riječi.

Keywords: digital lexicography, deep learning, representation learning

Ključne riječi: digitalna leksikografija, duboko učenje, učenje prikaza