

Understanding and producing words with high-school maths

Professor R. Harald Baayen, Quantitative Linguistics, University of Tübingen, explains how we can understand and produce words with high-school maths

Linear Discriminative Learning (LDL) is a computational theory of how speakers produce and listeners understand words. LDL is developed with the aim of providing a functional characterisation of the cognitive skills that allow speakers to express their thoughts in words, and that allow listeners to decode the intended message from these words. In psychology and cognitive science, the organisation of words in standard dictionaries has long served as a model for the organisation of lexical knowledge in human cognition.

Accordingly, in comprehension a word's form has to be identified first. This form then provides access to the corresponding meaning. From a machine learning perspective, the first step is a straightforward classification task: Given an audio signal or a visual pattern, one specific entry in a fixed list of words has to be selected.

A fixed word list can be made to work for English, as this language makes use of a very limited set of inflectional variants (e.g., hand, hands; walk, walks, walked, walking). For richly inflecting languages such as Estonian, where nouns have no less than 14 case forms in the singular, and another 14 case forms in the plural, or Turkish, a language with verbs that have more than 400 inflectional variants, fixed word lists with millions of entries become computationally unattractive.

Fortunately, the task of word recognition can be reconceptualised as a multi-label classification problem. An Estonian genitive plural is no longer seen as an atomic entry to be discriminated from millions of other such atomic entries. Rather, the classifier now has to predict three labels, one for the content word, one for singular number, and one for genitive case. LDL goes one step further, and replaces these labels, mathematically equivalent to binary vectors with one bit on and all others off, by real-valued high-dimensional vectors extracted from text corpora, building on methods from computational semantics.

For the Estonian genitive plural of the noun for 'leg', the vector for 'leg', the vector for plurality, and the vector for the genitive, are summed to obtain the semantic vector for the inflected noun. LDL also represents words' forms by numeric vectors. In the simplest set-up, these numeric vectors specify which letter or phone n-grams are present in a word.

For written words, low-level visual features have been used successfully. For spoken word recognition, LDL employs binary-valued vectors coding for features that are inspired by the tonotopy of the basilar membrane in the cochlea. A given feature specifies, for a specific frequency band to which a section of the basilar membrane is tuned, the pattern of change in amplitude over time.

It turns out that, surprisingly, straightforward methods from linear algebra, basically the high school mathematics for solving a set of linear equations, suffice to map words' form vectors onto their corresponding semantic vectors (comprehension), and to map words' semantic vectors onto their form vectors (production), see Figure 1. When phone triplets are used to construct form vectors, model performance is excellent with accuracies for production and comprehension around 95% or higher across languages as diverse as English, Latin, Hebrew, Estonian, and Russian. When used to model speech production, the kind of errors typically observed in human speech are also made by the model. The model also successfully predicts a range of empirical findings concerning human lexical processing.

Although LDL is developed with the specific purpose of increasing our understanding of human language comprehension and speech production, it may have potential for practical applications. For instance, for speech processing, current deep learning systems heavily depend on surrounding words to guess which words are encoded in the speech signal.

Mozilla Deep Speech, a state-of-the-art speech recognition system that is trained on thousands of hours of speech, performs very well for running speech, but without context, its performance collapses. Tested on the

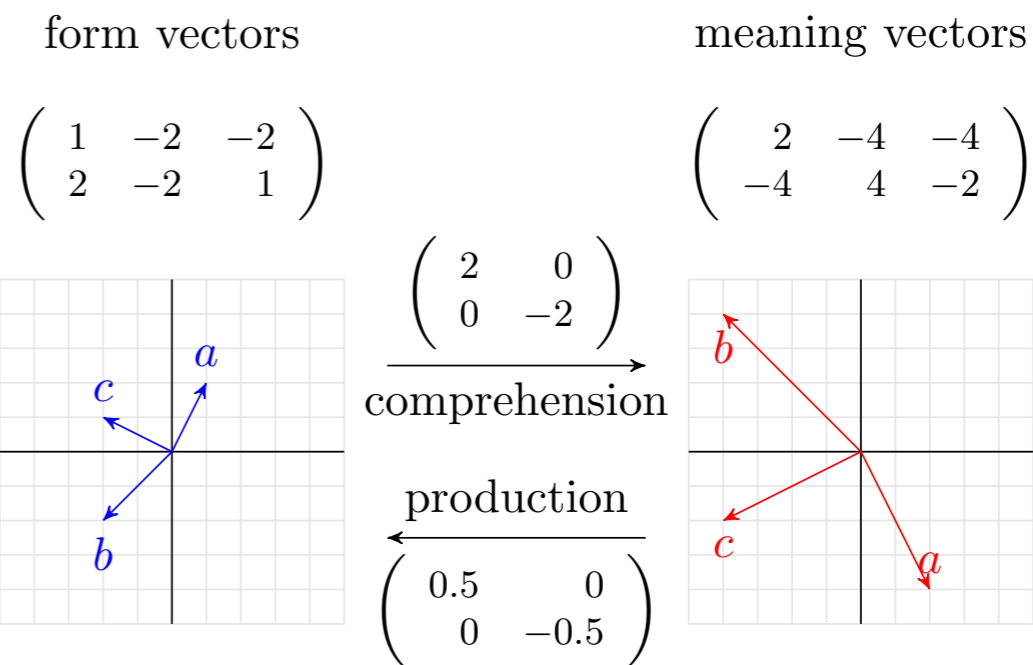


Figure 1: Word comprehension is modeled by matrix multiplication mapping form vectors onto meaning vectors, and word production is modeled by mapping meaning vectors onto form vectors.

audio of words excerpted from American TV news broadcasts, it recognised only 6% of the words presented to it in isolation. An LDL model trained on the audio of 10 hours of these news broadcasts performed with an accuracy of 8% (under 10-fold cross-validation). A cognitively motivated extension of the model resulted in an accuracy of around 13% (again under 10-fold cross-validation), substantially outperforming Mozilla Deep Speech.

Thus, LDL may prove useful as part of a speech recognition system for situations where context is unavailable. A major and as yet unresolved challenge for LDL is to get it to work not only for isolated words (as found in connected speech), but also for streaming connected speech. If LDL can indeed be extended in this way, it will provide an algorithm the carbon footprint of which is tiny compared to the unsustainably large carbon footprint of typical current AI solutions (see here at [MIT Technology Review](#)).

Another possible application is the assessment of the meaning of novel words, specifically, names for new inventions, technologies, and brands. Because LDL takes as input low-level visual or auditory features, it is able to project novel words into a high-dimensional space representing the meanings of existing words. A novel word's location in this space can then be used to evaluate what existing meanings it is most similar to, and hence what kind of associations and emotions it evokes.

Recent modelling results indicate that these associations are themselves predictive for both the speed of word recognition, and for how exactly speakers articulate novel words. These findings suggest that the model's predictions are precise enough for probing the semantics evoked by, for instance, novel brand names.⁽²⁾

Funding: European Research Council, project WIDE # 742545, principal investigator

R. Harald Baayen.

Website: (1) <https://www.quantling.org/ERC-WIDE/>
 Key publication: (2) Baayen, R. H., Chuang, Y. Y., Shafaei-Bajestan, E., and Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, 2019, 1(39), URL: <https://www.hindawi.com/journals/complexity/2019/4895891/3>



R. Harald Baayen
 Professor of Quantitative Linguistics
 University of Tübingen
 Tel: +49 (0)707 1297 3117
harald.baayen@uni-tuebingen.de
<https://www.quantling.org/ERC-WIDE/>
<https://uni-tuebingen.de/en/faculties/faculty-of-humanities/departments/modern-languages/department-of-linguistics/chairs/quantitative-linguistics/>