Software interfaces for teaching and learning in the fields of computational linguistics and natural language processing (NLP)

In the context of an awarded Teaching Development Grant from the University of Hong Kong, we have developed a number of visual interfaces to facilitate the understanding of various concepts and tools in the field of computational linguistics, and in particular in the subfields of natural language processing and text mining. These teaching and learning tools can be used for class activities as well as for assessments. Many of them are available in three languages: English, French and Mandarin.

You can find below a list of the interfaces, which are all available online (on a ShinyApp server). Some of them are pretty self-explanatory and/or contain a short description, other comes with descriptions in separate files that can be downloaded.

1. **TAME: Text Analytics Made easy**

This is the most-developed interface developed with the Teaching Development Grant. It allows users to choose between a number of text corpora (novels, news articles, song lyrics, plot summaries of movies from different country and of different genres) and conduct a number of text-mining analyses: term frequencies, co-occurrences of terms, collocations of terms, sentiment analysis, topic modelling.

The interface can be found here: https://keruiduo.shinyapps.io/text_mining

A short description is available here.

2. **An interface to study the diffusion of a(n) (linguistic) innovation**

This interface gives access to a simulation of the diffusion of an innovation (primarily a linguistic one, but the model is actually generic) on various social structures.

The interface can be found here: https://keruiduo.shinyapps.io/linguistic_diffusion

3. **A model of the competition between two linguistic variants**

This interface gives access to a computational model of the competition between two linguistic variants on a social network. It aims to answer questions such as: under which condition may a new variant spread to an entire community and replace a previous form? It is inspired by research papers published by Daniel Nettle in 1999 (here) or Ke et al. in 2008 (here).

The complete description of a teaching and learning activity with the platform can be found here.

The interface can be found here: https://keruiduo.shinyapps.io/competition_linguistic_variants

4. **An interface to experiment with morphosyntactic parsing**

This interface offers the possibility to investigate outputs of morphosyntactic parsing. One just needs to type a sentence, either in English, French, Mandarin with simplified characters or Mandarin with traditional characters, select the corresponding model in a menu, then start the parsing.

The interface then displays both a table containing the descriptions of the words of the sentence and a graphical representation of the sentence with the dependency relations between its elements.

The interface relies on the functions of the **udpipe** package in R (Straka et al., 2016).

The interface can be found here: https://keruiduo.shinyapps.io/morphosyntactic_parsing

5. **An interface to study the naming game**

Since their introduction by Luc Steels, naming games have enjoyed a wide success in computational approaches to the emergence of language. This interface gives access to a simple model of naming game structured around a few objects and weighted associations of objects and word forms in speakers' minds.

Two introductory papers by Luc Steels can be accessed here and here.

A detailed description of the model implemented is available here.

Users can modify the number of agents, the number of concepts and the probabilities to learn or create a word, and run a simulation for a given number of steps.

The interface can be found here: https://keruiduo.shinyapps.io/naming_game

6. **An interface to study the Levenshtein distance**

The Levenshtein distance is a well-known lexicographic distance for word forms. It corresponds to the number of transformations (addition, deletion, substitution) to shift from a word form to another.

The user simply needs to enter two word forms to see the corresponding distance, as well as the matrix used to compute this distance according to Wagner & Fischer's algorithm.

The interface can be found here: https://keruiduo.shinyapps.io/levenshtein_distance

7. **An interface to visualize word embedding**

Word embedding has garnered quite a lot of attention since its introduction in natural language processing: words can be turned into numerical vectors based on their collocations in a large corpus, leading to a dense semantic space where a number of operations can be performed – a clear improvement on one-hot encoding. A reference paper by Mikolov et al. (2013) can be found here.

With this interface based on a large set of English words and their embedding, the user can enter a word and observe its closest neighbors in the semantic space with a projection in a 2D space

The interface can be found here: https://keruiduo.shinyapps.io/word_embedding_graph

8. **An interface to study properties of word embedding**

A classic example aimed to highlight the semantic properties of word embedding is 'king – male + female = queen': one can subtract the 'male' property of 'king' and replace it by a 'female' property to find 'queen'.

Relying on a large dataset of numerical vectors (word embedding) for English words, this interface lets the user enter three English words A, B C, of their choice to look for the closest word to the vector result of the computation A – B + C. The previous example can easily be reduplicated, and other options be studied.

The interface can be found here: https://keruiduo.shinyapps.io/word_embedding

9. **An interface to illustrate Lotka-Volterra equations**

The Lotka-Volterra equations are a well-known set of differential equations to model the interaction between a predator and a prey.

This interface lets the user experiment with different initial numbers of preys and predators. They can also modify the values of the key factors of the equations, namely the prey intrinsic growth rate, predation efficiency, predator death rate and conversion efficiency. Different graphical outputs are available for inspection.

The interface can be found here: https://keruiduo.shinyapps.io/lotka-volterra

Code of the interfaces

All the interfaces have been developed in R with the shiny package

The code of all the interfaces can be found here.