What's in a word? Exploring words and their usage in the *Dictionnaire Vivant de la Langue Française*

Clovis Gladstone
clovisgladstone@gmail.com
ARTFL Project, University of Chicago
United States of America

Charles Cooney
chu.cooney@gmail.com
ARTFL Project, University of Chicago
United States of America

Tim Allen
timothy.d.allen@gmail.com
ARTFL Project, University of Chicago
United States of America

Originally funded by a startup grant from the National Endowment for the Humanities, *Le Dictionnaire Vivant de la Langue Française* ([DVLF)](#) is an experimental approach to dictionary compilation that aims to push the boundaries of what typical dictionaries offer. It is being developed as an interactive, community-oriented alternative to traditional methods of French lexicography, bringing together user-submitted definitions and definitions from standard French dictionaries (multiple editions of the *Dictionnaire de l'Académie Française, Dictionnaire de Littré…*), synonyms, pronunciation, as well as a wealth of information on word usage through the use of various computational methods. Our poster will show the DVLF's functionality, as well as demonstrate its originality.

Fundamentally, the most important aspect of the DVLF is that it aims to create an environment in which its user community rates, critiques, and adds to the collection's resources as it sees fit. As such, we have made a conscious effort to facilitate community engagement by allowing users to contribute to most sections of our website: definitions, usage examples, synonyms and antonyms. To date, users have added or submitted definitions/examples for hundreds of words, including **alexithymie**, **digiscopie**, **foucade**, and **nyctalope**. The new version of our site (released in January 2017) has been further enhanced to enable such engagement thanks to more visible links inviting users to contribute content, as well as a new responsive interface designed to work equally well on mobile devices and desktop computers.

In order to attract and engage the largest possible user community, the DVLF is entirely free and open source, and requires no registration. Internet users at large can access the site's resources and contribute material. The DVLF thus tries to adapt to changing word usage and incorporate neologisms so that users will be able to select the word senses and usage examples that they feel are most consistent with contemporary usage. In this manner, the DVLF mirrors the social and evolving nature of language, expanding upon the dictionary's traditionally normative role by giving French speakers and learners access to lexicographic tools so that they might interact with the evolving meaning of words and determine their own understanding of the language.

Over the last year, ARTFL has been working to develop and improve the DVLF. We noticed from statistics gathered from Google Analytics that our community of users was very diverse, accessing our website from Morocco, Tunisia, Canada, and many other francophone and non-francophone countries. This led us to focus a large part of our effort on diversifying content, including usage examples from a wider range of francophone sources in order to provide more coherent descriptions of emerging word usage from Francophone communities around the world, and therefore attract a larger global user base.

Over the course of this redevelopment effort, we decided to rewrite the entire codebase given the various reliability and performance issues we had experienced, switching from a Python-only infrastructure to a Go/Javascript environment. We have also worked to add additional contextual information to our content, starting with the most frequent collocates of any given word based a corpus of over 7,000 texts of French literature from the Middle-Ages to the late 20th century. Additionally, following the lead of recent work in computational linguistics and natural language processing, we also now provide the nearest neighbors based on the computation of word vectors using a word-embedding technique called Swivel. While we could have used alternative algorithms such as word2vec or Glove, we decided to use Swivel because it does not solely rely

only on word co-occurrence to construct vectors. The benefit of Swivel's approach is that it can yield similar words that do not actually occur together in the corpus (see Shazeer et al, 2016), which we thought was a valuable approach given that our dictionaries span across multiple centuries. To generate this word vector model, we used the same 7,000 texts and then extracted the top 20 words for every headword in our dictionary index, which is now displayed on the new version of our site.
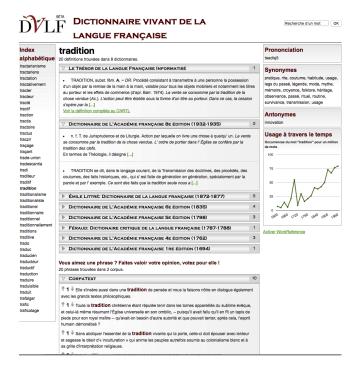
Appendix: Screenshots

Figure 2. New version of the site (went live on January 2017)

Figure 1. Old version of the site (retired on January 2017)