Re-imagining French Lexicography:
The *Dictionnaire vivant de la langue française*

Timothy Allen
Robert Morrissey
Glenn Roe

Abstract

The *Dictionnaire vivant de la langue française* (DVLF), developed by The ARTFL Project at the University of Chicago, represents an experimental, interactive, and community-based approach to French lexicography. The DVLF enables broad public access to a wide variety of linguistic tools and resources, with the goal of changing user interaction with dictionaries and providing better descriptions of emergent word use. In this article we describe the history of the DVLF and provide a survey of similar community-oriented electronic dictionaries. We then proceed to a presentation of the dictionary's many features, including the variety of its definitions and mechanisms for user interaction. The article concludes with a discussion of ARTFL's plans for the future developement of the DVLF.

Introduction

The following paper describes a recent lexicographical endeavor developed by the University of Chicago's ARTFL Project (American and French Research on the Treasury of the French Language) with the generous support of the National Endowment for the Humanities (NEH) Digital Humanities Start-up Grant Program. Currently in its "beta" testing mode, this project—*Le Dictionnaire Vivant de la Langue Française* (DVLF)—is an experimental approach to dictionary compilation that aims to offer an interactive and community-oriented alternative to traditional methods of French lexicography.

We see the DVLF as a paradigm-shifting dictionary in the sphere of traditional French lexicography. In the past half-century, conventional dictionary compilers have progressively taken advantage of computational methods to collect concordances and analyze text corpora, but we have yet to see a dictionary move beyond these relatively simple counting functions and use the computer for the lexicographical tool that it can be: a constructive social communication device for bringing together the diverse language communities that use dictionaries. While dictionaries might not typically be considered a type of "social communication device," we believe that opening the development of the dictionary to its users allows it to fully embrace its role as an inherently social object. Dictionaries, as descriptive linguistic resources, are generally meant to reflect a language as it is actually used among its speakers. Unfortunately, however, traditional dictionaries' descriptions of language have always been incomplete due to the limits of length and weight placed on printed books: dictionary editors have always been forced to pick and choose. Additionally, the relatively small set of lexicographers who compile dictionaries could never be expected to track in an exhaustive manner the continually evolving state of a given language. The act of providing speakers with the tools to document their own descriptions of the language, however, suggests a different model for what a dictionary can be. Dictionaries such as the DVLF allow anyone with access to the internet the ability to document their language as they speak it, and consequently offer users a timely, wide-ranging understanding of language as it exists in that moment.

While the DVLF incorporates many traditional dictionary sources among its lexicographical resources, it also creates a virtual environment in which its user community has the ability to rate, critique, and add to those resources as it sees fit. Perhaps most notably, users have the ability to add their own words, definitions, and usage examples to the lexicon. These technical innovations, along with the fact that the site is entirely free and open source, have been implemented in order to attract and incorporate the largest possible user community. Internet users at large are able to access the site's resources and contribute material. Thus, the DVLF will be able to adapt swiftly to changing word usage and quickly 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 evolving nature of language, expanding upon the French dictionary's traditionally normative role by giving French speakers and learners access to lexicographic tools and enabling them to interact with the changing meaning of words that, ideally,

will help determine their own understanding of the language. The DVLF can thus be thought of as a dynamic, adaptive lexicographic search engine, informed by, but standing in stark contrast to existing static reference works.

New concepts of lexicography

In 1957 the French government, under the guidance of lexicographers Paul Imbs and Bernard Quémada, initiated the creation of a new dictionary of the French language, the *Trésor de la Langue Française* (TLF).[1] In order to provide lexicographers access to a large body of word usage examples, the editors of the TLF decided to transcribe an extensive selection of French texts for use with computers, specifically "data processing machines and punch cards." The idea of involving computers with lexicography was a radical one at the time, and twenty years later, a corpus totaling some 150 million words had been created, representing a broad range of written French—from fiction and poetry to biology and mathematics—stretching from the sixteenth to the twentieth centuries.

It soon became apparent that this corpus of French texts was an important resource not only for lexicographers, but also for humanists and social scientists engaged in French studies across the disciplines, and on both sides of the Atlantic. The product of this realization was the Project for American and French Research on the Treasury of the French Language (ARTFL), a cooperative enterprise established in 1981 by the *Centre National de la Recherche Scientifique* and the University of Chicago. Over the years, ARTFL has steadily expanded its corpus of French texts to the point that it now contains some 168 million words in nearly 3,000 works.

From its inception, then, the ARTFL Project may be understood as the inheritor of a once-revolutionary approach to dictionary compilation. Today, however, no modern dictionary would be taken seriously if it did not have a corpus of natural language at its foundation, and thus the TLF *informatisé* (TLFi)[2]—the 2002 electronic version of the originally-hardbound TLF, which was completed in 1994[3]—does not strike many as a particularly ground-breaking dictionary. And, although the ARTFL Project continues to maintain an extensive collection of electronic dic-

---

[1] Pruvost 2006, 86**.**

[2] *Le Trésor de la Langue Française informatisé*: http://atilf.atilf.fr/tlf.htm

[3] Pierrel 2004.

tionaries in many languages, none of these might rightly be called innovative. These dictionaries, despite their various forms and disparate editorial histories, are all products of traditional lexicographic methodology, one in which trained lexicographers determine a word's meaning, history, and proper usage. This is not to say that we wish to reject the editorial work of trained lexicographers. Indeed, the DVLF includes entries from a number of traditional dictionaries so that contributors might more readily recognize the site as a serious resource grounded in lexicographical standards. We feel that these dictionaries can serve as models for our contributors and consequently keep the DVLF from mirroring the limitations of many other community-oriented electronic dictionaries. To date, experiments in alternative dictionary compilation have met with varying degrees of success, and perhaps a few of these projects deserve mention.

The "Urban Dictionary,"[4] an online dictionary of non-standard English usage written by users, offers a glimpse of what a community-based dictionary could be, yet the site is an anarchic collection of nonce words and arbitrary example sentences that appear to have been entered solely for the entertainment of the writer and his circle of friends. "*Le Dico des mots imaginaires*,"[5] a similarly user-defined dictionary for French, appears to have an insubstantial user community and no search functionality. The notion of "imaginary words" would also seem to call into question the general, quotidian utility of the site. "*Le Dictionnaire de la Zone*"[6] is a French dictionary to which anyone can contribute, yet it limits itself to the urban slang of the "banlieue." Finally, "*Bob: Dictionnaire arg. pop. fam.*,"[7] an online French slang dictionary, has at its base a wiki to which anyone, presumably, can contribute linguistic material, yet it is not clear that this wiki feeds into actual dictionary entries. Furthermore, none of these resources provide substantial usage examples from external corpora.

The sites mentioned above display a very limited range of information in their focus on presenting non-standard usage. Slang dictionaries have long existed and, while they may provide a fascinating perspective on language, they are not particularly innovative. One online dictionary that may be said to display a truly innovative spirit, however, is

---

[4]Urban Dictionary: http://www.urbandictionary.com/.
[5]*Le dictionnaire des mots imaginaires*: http://dico-des-mots.com/.
[6]*Le Dictionnaire de la Zone*: http://www.dictionnairedelazone.fr/.
[7]*Bob, Dictionnaire arg. pop. fam.*: http://www.languefrancaise.net/bob/

Erin McKean's recent project Wordnik.[8] Wordnik's goal is to become a home for "all the words," not just the subsets of words found in typical dictionaries. The site combines usage examples from a growing corpus of English sentences with Twitter search results; Flickr photos; audio pronunciations provided by users and licensed from other commercial dictionaries; corpus analysis tools; user-generated annotations; and synonyms, etymologies, and definitions from traditional dictionary sources to provide "as much information as possible...for every word in English."[9] At the same time, Wordnik's entries are highly structured and neatly organized. Despite, or perhaps because of, its breadth and inclusiveness, Wordnik is clearly a serious lexicographical tool.

The theory underlying the presentation of this battery of lexicographical data is the notion that today's electronic tools have made the traditional paper dictionary obsolete. Whereas bound paper dictionaries are limited both by the physical constraints and the economic concerns of their publishers—and so must decide which words are "allowed" into the dictionary—Wordnik can take advantage of the sprawl of the internet to supply its users with a near-endless stream of information about a particular word. Able to disregard the strictures of traditional dictionaries, Wordnik does away with the idea of the dictionary as the "traffic cop" of language; antiquated, non-standard, and technical terms are displayed with equal prominence alongside the mainstream lexicon. Wordnik "expands the dictionary" insofar as it gives its users the data and tools they need to comprehend more fully a word's variant usage and meaning.

At its most general level, the DVLF can be thought of as an online dictionary for French that follows in the spirit of the English-language Wordnik. The DVLF is designed to give French speakers from around the globe the ability to interact with and share linguistic information for all domains of French usage. As a result of the practically unlimited expandability of the DVLF, and following Ledegen's recent assertions, *"[I]l devient en effet possible d'ouvrir sur les autres domaines où la variation se manifeste, autres domaines qui sont indissociables de celui du lexique, et de mettre à la disposition de l'utilisateur des analyses des rares traits morpho-syntaxiques qui trouvent difficilement leur place dans les dictionnaires, mais particulièrement fortement les usages.*" [In effect it becomes possible to open onto other domains in which variation manifests itself, other domains that are inseparable from that of the lexicon, and that put at the disposal of the user analyses

---

[8]http://www.wordnik.com/.
[9]http://www.wordnik.com/about.

of rare morpho-syntaxic traits that have great difficulty finding their way into dictionaries, but that strongly represent *usage.*][10] Indeed, in addition to building an expansive array of lexicographical tools and resources in the vein of Wordnik, we have also developed a means for our users to rate and immediately change the display of items in our usage example corpus. This rating feature, which is described in more detail below, will soon be expanded to include definitions, as well. As DeSchryver notes, there is still a question as to whether "average users" will be "able to process the raw data and to turn these into lexicographically sound information," but that "the technology is there, and advanced users might benefit from it."[11] We are convinced that this tool, combined with the DVLF definition submission feature and a broad selection of linguistic utilities, will inspire our user community, consisting of both average and advanced users, to work to improve the range and quality of dictionary entries and also to provide more coherent descriptions of emergent usage from Francophone communities around the world. In so doing, DVLF users will move language beyond the restrictions of traditional lexicography, expanding the functionality of the dictionary to an extent rarely before seen, particularly in relation to French dictionaries. As described by Corbin and Gasiglia, *"[L]a lexicographie française commerciale, même maintenant qu'elle n'est plus démunie de ressources informatisées, tend à mettre davantage l'accent, dans ses discours publicitaires, sur l'ancrage culturel des répertoires que sur les technologies qu'elle utilise ou les concepts linguistiques qu'elle met en oeuvre."* [Commercial French lexicography, even though no longer devoid of computerized resources, tends to place greater emphasis, in its advertising discourse, on the cultural anchoring of its collections rather than on the technologies that it uses or the linguistic concepts that it implements.][12] In the view of these French lexicographers, French lexicography is generally a resolutely conservative field. Corbin is even cited as speaking of "*régression*" [regression] on the part of French dictionary compilers who reject the useful contributions of modern linguistics to lexicography.[13] Moreover, Gasiglia takes on a rather defeated tone when she writes that corpora and corpus exploration tools "*ne semblent pas avoir beaucoup d'avenir en France*" [seemingly have little future in France][14] and that "*ce ne sont pas les corpus qui constitueront prochainement les meilleures*

---

[10]Ledegen 2008, 273, emphasis added.
[11]De Schryver 2003, 157.
[12]Corbin and Gasiglia 2009, 20.
[13]Béjoint 2009, 119.
[14]Gasiglia 2009, 237.

sources d'information chez les éditeurs français" [it is not the corpora that will soon constitute the best sources of information for French editors].[15] The situation of contemporary French lexicography as described by Corbin and Gasiglia is rather surprising given that, as discussed by De Schryver, user access to the lexicographer's corpus has been a dream for many over the past two decades: "Since the early-1990s, various scholars have expressed the wish to offer users of EDs [Electronic Dictionaries] the same wealth of information lexicographers find in corpora, in other words, to include a corpus cum query tools as integral parts of an ED."[16]

It is our belief that the DVLF, on the heels of such enterprises as Wordnik, is well on its way to fulfilling this "lexicographer's dream." Indeed, the example corpus, discussed in greater detail below, lies at the heart of the resources that the DVLF offers to its users, and its use within the framework of the site is by and large unheard of in the "regressive"[17] realm of French lexicography.

Research and development

As described above, the ARTFL Project was first conceived as a means to disseminate a digital collection of French works that was also the cornerstone of one of the world's first corpus-based dictionaries. From these roots in French lexicography, the ARTFL Project has grown to become one of the primary sources for French language analysis in the United States. In addition to its ongoing creation and maintenance of natural language databases, ARTFL has developed several open source text analysis tools and is actively engaged in the exploration and implementation of new techniques that bring the power of data mining and machine learning techniques to bear on humanities and social science research.

Among ARTFL's many databases are a number of historical French dictionaries, many of which currently exist as constituents of the *Dictionnaires d'autrefois* collection (DAF).[18] The DAF represents only a fraction of ARTFL's dictionary and language reference experience, however, as it also maintains a publicly accessible collection of English reference works,

---

[15]Gasiglia 2009, 241.
[16]De Schryver 2003, 167.
[17]Corbin 2002. Cited in Béjoint 2009, 119.
[18]*Dictionnaires d'autrefois*: http://artfl-project.uchicago.edu/content/dictionnaires-dautre-fois.

a French-English bilingual dictionary, a French verb conjugator/reverse conjugator, Diderot and d'Alembert's seminal *Encyclopédie*, and dozens of South Asian dictionaries. This bevy of free language resources has, over many years, developed a sizable user community that we expect will engage with the DVLF's interactive features and shape the core of our approach. Perhaps the most important aspect in cultivating our potential user community will be fostering a large enough Francophone user base to guarantee an effective level of native-speaker interaction with the DVLF. Through ARTFL's long-standing relationships with *Analyse et traitement informatique de la langue française* (ATILF) of the CNRS, *L'École des chartes, Université de Paris IV (Sorbonne), le Groupe Fabula,* and the University of Ottawa's *Modeling Change: the Paths of French* project, we have released the DVLF to a large and diverse community of Francophone teachers and speakers, all of whom we anticipate will engage with its interactive features and inform its initial release.

Of course, ARTFL is not solely a proprietor of electronic dictionaries. ARTFL has an extensive history of working with large full-text corpora, most importantly FRANTEXT, which served as the usage example base for the TLFi mentioned above. In order to retrieve and analyze these corpora, ARTFL, in collaboration with the Digital Library Development Center at the University of Chicago, developed PhiloLogic,[19] a robust, customizable search engine that incorporates traditional notions of philology and applies them to large collections of texts. It would be inappropriate to enumerate all of PhiloLogic's features in this space, yet there are a handful of features particularly germane to the DVLF that warrant further discussion.

At its core, PhiloLogic builds concordances. PhiloLogic's concordances provide users with information about a word's usage in various contexts, allowing them to track a word's frequency of use over time, and display the frequency of a word's collocates. These sorts of features currently manifest themselves in the DVLF in usage example sentences and word frequency graphs, and will allow us eventually to provide word collocation clouds.

Taking a cue from Michael Halliday in the field of functional linguistics, ARTFL developed PhiloLogic with the ability to generate "theme-rheme" reports for a given word. In other words, PhiloLogic can look at the concordance results for a particular word and identify when that word is "clause-initial" or in any other generally-defined position in a sentence.

---

[19]PhiloLogic: http://sites.google.com/site/philologic3/.

The functional linguistic notion that informs this feature is the idea that, if a word occurs initially in a sentence or clause, the sentence's writer means to mark that word as the "subject" or "topic," and thus as particularly important to the idea being expressed.[20] The "theme-rheme" report feature was originally implemented to allow ARTFL's users to track the relative importance of particular word usage in various authors and over time. In the case of the DVLF, however, we believe that we can employ the "theme-rheme" function to aid in the automatic identification of example sentences that are particularly illustrative of a word's meaning, hypothesizing that a word's clause position has a direct relation to that word's relative significance.

In recent years, the ARTFL Project has undertaken work in the realm of machine learning and data mining, leading notably to the development of two extensions to PhiloLogic: PhiloMine and PhiloLine. These programs allow users of PhiloLogic databases to locate related works within and between databases through the identification of similar passages and other mesures of textual similarity. FRANTEXT users are currently able to use these tools to search for similar passages in the database, and the ARTFL Project is continually working to apply data mining techniques to its various other databases. As outlined below, we expect to leverage our significant technical capacity for text analysis and our experience with the construction of machine learning and data mining applications in the further development of the DVLF's word usage database. These applications will be continually refined over subsequent development phases under the aegis of ARTFL's research and development staff.

DVLF Features

During the DVLF's first development phase, we undertook to build back-end databases and a user interface (UI) for management and display of definitions from our currently-held historical dictionary resources, entries from the modern TLFi, and usage examples from our corpus of French texts. These components form the core of the lexicographical content that the DVLF displays to its users.

The task of the UI is primarily one of combining the considerable lexicographical resources that ARTFL already possesses and turning them

---

[20]Halliday 2002, 200. This notion holds across languages. Betsy J. Kerr states, "In English and French, as in many languages, there is a tendency to place the topic in sentence-initial position, for example in subject position." (Kerr 2011, 924.)

into a cohesive resource that our user community can easily navigate. There are essentially three sets of information that the DVLF displays in response to a user query: dictionary definitions, usage example sentences, and linguistic tools like thesauri and pronunciation information. Considering the depth of ARTFL's resources, these three items represent a potentially massive amount of text, much more than could be usefully displayed at once on a single web page. Thus, we have created a "summary page" for each DVLF entry that is divided into three columns. Two narrow columns contain an alphabetical index (or "word wheel") for the user's reference and our ancillary linguistic tools, while the wider center column contains all of our available definitions, as well as a selection of example sentences from our corpus. Features that we expect to implement soon, such as our verb conjugator and word collocation cloud, will most likely be displayed in a footer section at the bottom of the summary page.

     The DVLF's definitions for a given word come from up to eight historical French dictionaries: Jean Nicot's *Thresor de la langue française* (1606), five editions of the *Dictionnaire de l'Académie française* (1694, 1762, 1798, 1835, 1932-1935), Féraud's *Dictionnaire critique de la langue française* (1787-1788), and Littré's *Dictionnaire de la langue française* (1872-1877). The DVLF also provides introductory definition snippets and a link to a word's entry at the website of the *Trésor de la Langue Française Informatisé.* All of these dictionary entries occasionally add up to a great deal of text (the word "*grand*" has 94 entries for example), and so they are each expandable and collapsible in the browser. The two most modern dictionaries (the TLFi and the 1935 edition of the *Dictionnaire de l'Académie française*) are displayed first and are "expanded" by default. Users can then choose to open additional entries if they so desire.

     Using data borrowed from the CRISCO laboratory at the University of Caen, we are also able to provide selections of synonyms and antonyms for more common vocabulary items.[21] Some of the thesaurus items have been ranked according to CRISCO's proprietary measurement of similarity, while the rest are presented alphabetically. When lists of synonyms grow larger than 25 items for a given word, we hide the trailing synonyms and give the user the option to expand the list.

     The DVLF's "thesaurus" occupies a narrow column on the right side of each word's summary page. This column also contains an IPA pronunciation key that we extracted from the TLFi and an embedded page

---

[21]http://www.crisco.unicaen.fr/des/.

from WordReference.com that provides translations for both individual words and idiomatic expressions. On the opposite side of the page is an alphabetical index of nearby dictionary entries that simulates the experience of browsing a printed dictionary page and helps users to find words related to their initial query.

If a given word exists in ARTFL's FRANTEXT database, the DVLF displays a graph of the frequency of the word's usage over time, as well. This corpus currently comprises 2,900 French texts and contains more than 168 million words. FRANTEXT is primarily a literary corpus, and so jargon and technical terms may be absent from it. FRANTEXT also does not contain many works that date later than the second half of the twentieth century. These caveats are important to bear in mind, but are mitigated, we believe, by the DVLF's use of additional corpora.

The DVLF employs two corpora from which usage examples are gathered. The first, fixed corpus is called Corpatext and is freely available for public use.[22] Corpatext contains fewer words (37 million) than FRANTEXT and associates very little metadata with its texts, but it is open source and avoids problems that could stem from the public use of FRANTEXT, which is technically licensed to ARTFL by ATILF and the CNRS. Corpatext includes a wide variety of texts and was very easily parsed into individual sentences. Up to ten sentences from Corpatext are displayed to DVLF users on every word summary page.

Our second usage example corpus was constructed by ARTFL for use in the DVLF. This corpus consists of hundreds of thousands of francophone webpages (including scientific journals, major French and Swiss newspapers, and Quebecois lifestyle websites) that were distilled down to their primary texts and then divided into sentences. This corpus contains nearly 11 million words and is designed to fill the void of modern usage examples left by Corpatext and FRANTEXT. We can also add data to the corpus at will, and are currently working to incorporate text from sources in Francophone Africa. Once again, a maximum of ten example sentences from this corpus is displayed to the user for a given word. These sentences, along with those from Corpatext, can each be "voted" up or down by DVLF users according to their perception of the sentence's usefulness or lexical representativeness. Votes are tallied to provide a score for each sentence, and sentences with high scores are displayed near the top of the list of example sentences, while those with low or negative scores can be dropped from the list entirely. This movement of sentences

---

[22]http://www.lexique.org/public/corpatext.php.

according to votes happens entirely automatically on the DVLF website. As we will discuss below, we hope to analyze these scores and their associated sentences in order to mine our corpora progressively for better example sentences. Currently, sentences are selected for display if the query word occurs within the first five words of the sentence. This method of sentence selection, though informed by ARTFL's experience with theme-rheme reports, is admittedly rather simple, and we look forward to fine-tuning it in our next development phase.

One final feature of the DVLF that deserves mention is the ability that it provides users to submit new material to the dictionary. Though the DVLF contains more than 125,000 headwords, it cannot come close to representing all of the words or word senses of the French language without input from users. In order to allow for this input, we have placed a link at the bottom of every word summary page that takes users to a page where they can submit additional information about a word, whether that information be a new definition, a new sense, or an example sentence. If a sought-after word does not have an entry in the DVLF, we provide a link in our "word not found" page to the user submission page. Any new content provided by users is added immediately to the DVLF's database and visible to the user via a simple page refresh.

Future Development

In the short term, we plan to carry out additional research into automatic selection of useful example sentences. We will leverage ARTFL's machine learning experience to analyze the example sentences chosen by the human editors of the Littré dictionary and the TLFi. We expect to discern patterns in these sentences that will then help us to identify lexically salient sentences in our corpora more broadly. Machine learning will allow us to merge corpus- and document-level metadata with information about the sentence itself, and to learn rules such as: "Good example sentences for words that are rare in the overall corpus tend to occur at the first sentence-initial usage of a word in a document that uses that word frequently." Sentences would automatically be assigned a score and the highest-scoring example sentences would be shown to the user first, allowing the user to grasp more quickly the important patterns of usage for that term.

In our second development phase, we will also focus on the analysis of user-generated content and the cultivation of our user community. Our first task will be to combine the ratings data that we collect from our users with the automatically-generated scores mentioned above. We will then use these combined scores to either promote or demote the exam-

ple sentences that are displayed to the user. In other words, highly-scoring example sentences will be more likely to be displayed to users at the top of our example list, while lower-scoring examples will be pushed to the bottom. In keeping with the notion that the DVLF is not a staid, traditional dictionary, however, we will continue to display unranked sentences near the top of our example list so that our users are always shown new examples of usage, and so that we will have a continuous stream of ranking data to apply to our sentence database. This stream of data will then be leveraged into a system that allows for automatic recalculation of a sentence's salience score in real time, thus creating a sentence database that is entirely fluid and constantly changing according to our user community's input. We believe this sort of cultivated example sentence database to be the first of its kind available in conjunction with a freely-accessible public dictionary.

In order to cultivate the engaged user community that we expect for this resource, the second development phase will also entail the implementation of a user management system. This system will allow DVLF users to log in to the site, access their account information and leave relevant content on the site. While any anonymous user will be able to access all of the DVLF's lexicographical information and rate sentences, users who create accounts will have the added ability to suggest related words and comment on example sentences and definitions. Users with DVLF accounts will also be able to track previous words they have looked up on the site, along with any content they may have added. In fostering this sort of personal engagement with the DVLF, we believe that we will attract a core of language enthusiasts on both sides of the Atlantic who will contribute to make the site more dynamic than traditional dictionaries and thus more valuable to our general users.

In later stages of development, we envisage further enhancements to both our users' experience with the DVLF and the contributions of information science research to the internal workings of the site. In terms of user experience, we imagine adding salient images to word summary pages, building a version of the site for mobile phones and tablet computers, and implementing more highly structured "thesaurus" entries. We would also like to incorporate audio pronunciations and, much like Wordnik, a means for users to record their own pronunciations. As for new information science research, we are currently developing a project in virtual normalization of historical French text. Normalized spellings will help us to provide users with a wider range of historical use cases and will also make it so that DVLF users can more easily find older definitions for some words that have undergone orthographic changes over time.

Conclusion

In her recent essay on the digital evolution of lexicography, French lexicographer Nathalie Gasiglia poses an important question about what she and other like-minded lexicographers imagine the dictionary could be: "*Serait-il envisageable, pour un éditeur, de proposer un dictionnaire électronique qui contiendrait 'toutes' les informations susceptibles d'être utiles et qui disposerait d'une interface dont la lisibilité permettrait aux utilisateurs de s'orienter aisément vers la sélection d'informations pertinentes pour répondre à ce qui motiverait chaque consultation* ?" [Would it be possible for an editor to offer an electronic dictionary that would contain "all" the information likely to be useful and that would also include an interface whose readability would permit users to move easily towards the selection of pertinent information that would respond to the questions posed by each consultation?][23]

The question is important despite, or perhaps because of, the very general terms in which it is posed. Gasiglia is asking for two deceptively simple things in a dictionary: all possible useful information delivered in a manner that is easily read and navigated by the dictionary user. Paper dictionaries have always been limited by physical constraints of size and weight, but electronic dictionaries do seem to have the capacity to fulfill the dream that Gasiglia, and others, describe. We believe that the DVLF, unfinished as it is and perhaps always must be, is unique within the sphere of French lexicography for the way in which it provides the French-speaking public a chance to use the lexicographer's tools at a remove from traditional dictionary sources. This large-scale experiment in community lexicography speaks with the authority of centuries of French lexicographical expertise but also allows all speakers of the French language to contribute their own, changing understandings of the language to the resource. With this in mind, we look forward to the ways in which our users dynamically guide the direction of the DVLF in the future.

References

Béjoint, Henri. 2009. Lexicographie et linguistique : quelques réflexions sur le domaine anglais. In *Changer les dictionnaires ?*, edited by Pierre Corbin and Nathalie Gasiglia, 117-158. Lille, France: Presses Universitaires du Septentrion.

---

[23]Gasiglia 2009, 267.

Corbin, Pierre. 2002. Lexicographie et linguistique : une articulation difficile. L'exemple du domaine français. In *De la lexicologie à la lexicographie / From Lexicology to Lexicography*, edited by F. Melka and M.C. Augusto, 8-37. Utrecht: Utrecht Institute of Linguistics OTS.

Corbin, Pierre and Gasiglia, Nathalie. 2009. Changer les dictionnaires ? Une pluralité d'approches. In *Changer les dictionnaires ?*, edited by Pierre Corbin and Nathalie Gasiglia, 7-38. Lille, France: Presses Universitaires du Septentrion.

De Schryver, Gilles-Maurice. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16.2: 143-199.

Gasiglia, Nathalie. 2009. Évolutions informatiques en lexicographie : ce qui a changé et ce qui pourrait émerger. In *Changer les dictionnaires ?*, edited by Pierre Corbin and Nathalie Gasiglia, 235-298. Lille, France: Presses Universitaires du Septentrion.

Halliday, M.A.K. 2002. Some lexicogrammatical features of the *Zero Population Growth* text. In *Linguistic Studies of Text and Discourse*, edited by Jonathan Webster, 197-227. London: Continuum.

Kerr, Betsy J. 2011. A Corpus-Based Study of Indefinite *des* and Generic *les*. *The French Review* 84.5: 918-933.

Ledegen, Gudrun. 2008. En Contrepoint : Autres expériences, autres regards. In *Le français des dictionnaires*, edited by Claudine Bavoux, 269-276. Brussels: Groupe De Boeck.

Pierrel, Jean-Marie. 2004. La préface du TLFi. http://www.inalf.cnrs.fr/atilf/produits/tlfi_preface_jmp.htm accessed 8/9/2011.

Pruvost, Jean. 2006. *Les dictionnaires français, outils d'une langue et d'une culture*. Paris: Ophrys.

Acknowledgements