

From Look-Ups to Learning: How Vocabulary.com Reinvents the Dictionary

The digital transformation of dictionaries provides fresh opportunities for lexicographers to engage with generations coming of age in the electronic era. Vocabulary.com has been at the forefront in making online dictionary resources maximally useful to a 21st-century audience. This paper examines how Vocabulary.com is pioneering digital lexicography, creating an unprecedented tool for language learning.

Look up a Word, Learn it Forever

Each “word page” in the Vocabulary.com Dictionary is built to achieve a simple goal: providing the most efficient path toward learning what a word means and how it is used in context. This goal is accomplished by moving beyond the traditional structure of dictionary entries to present engaging text that explains word usage, as well as commandeering examples from a massive textual corpus to illustrate words in naturally occurring contexts. Furthermore, by integrating these rich lexicographical resources into an adaptive learning program, word-learning becomes an active, dynamic process. Words come alive for language learners, rather than being dry, static entities in a dictionary entry or on a flash card.

By designing its online dictionary with language learners in mind, Vocabulary.com brings together the latest insights from electronic corpus-based lexicography with the science of learning. The speedy and intuitive online interface further eases the word-learning process. Thus, with the Vocabulary.com Dictionary, looking up a word constitutes the first step on a path to lifelong mastery of the word’s meaning and usage.

To understand how Vocabulary.com has reimagined the dictionary as a dynamic tool for language learning, we will examine its lexicographical innovations in detail. In her introduction to the edited volume *Electronic Lexicography*, Sylviane Granger lays out several areas of innovation that lexicographers are currently exploring in online reference, including *hybridization, more and better data, efficiency of access, customization, and corpus integration* (Granger 2012). By excelling in each of these areas, the Vocabulary.com Dictionary positions itself as a word-learning resource of unrivaled scope and efficiency.

Fusing the Dictionary with a Learning Program

In an online environment, it is much easier to combine different types of reference resources, which Granger (2012) terms *hybridization*. A typical example of hybridization would be the combination of a dictionary and a thesaurus, and the Vocabulary.com Dictionary is fully integrated with Thinkmap's award-winning Visual Thesaurus; from any word page in the dictionary, the user can view an interactive word map encouraging the exploration of words and meanings in an engaging, dynamic interface (more on this below). But Vocabulary.com also breaks new ground by integrating its dictionary with an adaptive learning program.

More than 10,000 words in the Vocabulary.com Dictionary have been made learnable: by simply clicking on the "learn" button, the word is immediately added to the queue of words that the user is learning through Vocabulary.com's gameplay, the "Challenge." For each learnable word, there are on average 10 multiple-choice questions that have been developed to cover the word's different meanings, using a variety of different question types. Thus, a word look-up need not be a transient moment, but can instead open a door to discovery, introducing the word to the user's permanent vocabulary.

Conversely, dictionary resources are fully built into the Vocabulary.com Challenge: for instance, if a user incorrectly answers a question, a helpful explanation of the word is immediately presented, and the rest of the word's page in the Vocabulary.com Dictionary becomes available for a deeper dive into semantics, usage, spelling, and pronunciation. This two-way integration ensures that the Challenge and the Dictionary work together in a mutually enriching fashion.

The Most Useful Word Data, Pleasingly Presented

Compared to their print counterparts, online dictionaries promise to deliver *more and better data*. The limitations of page extent are no longer a concern in the online space. Breaking the "straitjacket of printed page and alphabetical order" (Atkins 1996:516) requires a thorough rethinking of what features might be most useful for a user, when an entire web page is available for explicating a particular lexical item. The danger then becomes overwhelming the user with a surfeit of information about a lexical item, which can become even more unwieldy if advertisements are also running on a web page.

Judicious use of web design requires knowing *which kinds* of language-related data (text, audio, or visual) will be of most use and how best to highlight them in an online environment. The Vocabulary.com Dictionary takes advantage of the electronic format by presenting the vocabulary learner with information that encourages a full understanding of

a word's meaning. Crucially, word pages in the Vocabulary.com Dictionary feature more conversational explanations of words in addition to typical dictionary definitions. This more discursive approach to defining words by means of "full-sentence definitions" has been explored in the past primarily in dictionaries targeting learners of English as a foreign language (Atkins and Rundell 2008:441). But the value of explaining words in an engaging style is just as important for native speakers of English who would not be able to appreciate the full range of how a word is typically used from a traditional dictionary definition. Researchers in how students engage with dictionaries have long called for "definitions that characterize a word's use and are phrased in language that makes meaning easily accessible to young learners" (McKeown 1991:154). The Vocabulary.com Dictionary takes exactly this approach to elucidate thousands of vocabulary words in a vivid manner that ensures that the explanations will be memorable.

Consider the word *quadruped*, which a standard dictionary would define simply as "an animal with four feet." In the Vocabulary.com Dictionary, the word page for *quadruped* leads with this light-hearted explanation:

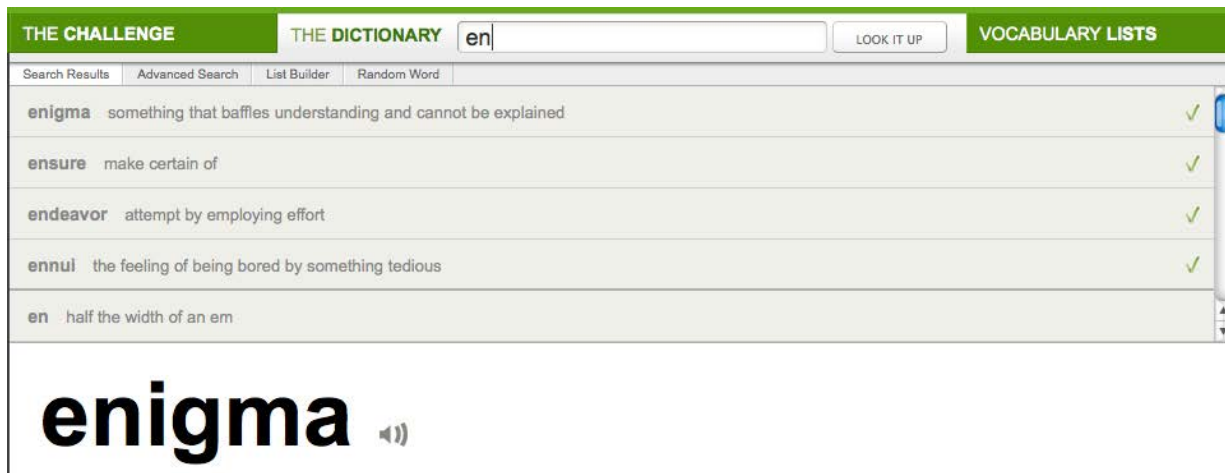
A squirrel, a zebra, a deer, a wolf, and a grizzly bear meet in a field. Yes, a disaster in the making, but also a bunch of *quadrupeds* — animals that walk on four feet.

The second part of the explanation goes into more detail, providing morphological and etymological background to help the meaning stick in the learner's mind, while still maintaining an easy-going, discursive tone:

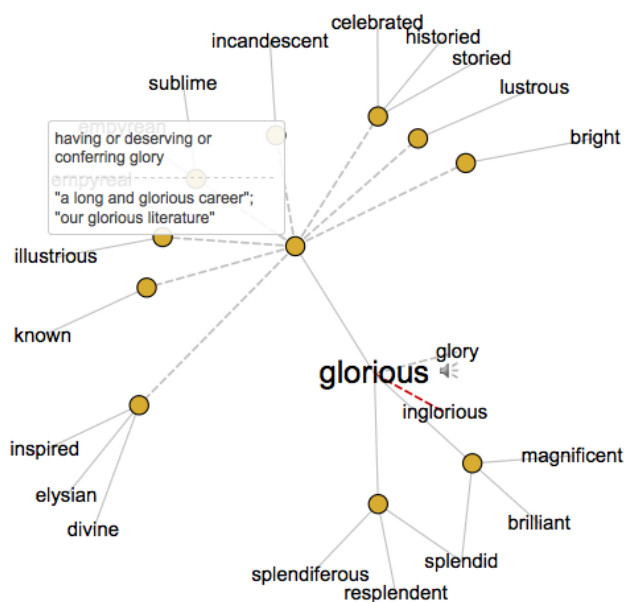
Cut *quadruped* in half and it makes sense: *quadru* means four, like when a woman births four babies they are called *quadruplets*. And *-ped* is for the feet: think of centipedes and millipedes, insects that have so many feet it's disturbing. A human is a biped because they walk on two feet. If you meet a human with four feet, you could call him a *quadruped*. You could also call the circus and let them know their quadruped is loose.

Along with these clear, approachable explanations, the word pages also include audio pronunciations, "word families" of inflectionally and derivationally related forms (indicating relative corpus frequency for each form), and corpus example sentences arranged by genre and topic. Additionally, if a word is part of the user's Vocabulary.com learning program, the dictionary page can graphically indicate progress toward mastery of that word. All of this information is presented in a clean, uncluttered design.

hyperlinked navigation and search functionality are deployed well, it can in fact be much easier to home in on the desired lexical item than it would be by paging through a printed work. Vocabulary.com uses autocomplete and autocorrect to provide “results as you type” and also corrects spelling errors by means of the Soundex phonetic algorithm. Instant results can be displayed from typing just a few letters.



As the Web excels as a visual medium, effective data visualization can be achieved in any number of creative ways, using different spatial metaphors. The Visual Thesaurus, which is integrated into the Vocabulary.com Dictionary, creates interactive displays of the relationships between words and between senses of words by means of elastic, spring-like graphs. Moving from node to node through this type of semantic visualization, the jumps can be unexpected, allowing for the emergence of a different kind of serendipity than the kind that a print reference normally affords.



Personalizing the Word-Learning Experience

The move from the static printed page to the dynamic web page encourages a responsive design that is tailored to a user's needs. This becomes a crucial concern in creating resources for English language learners. As Granger (2012) suggests, *customization* in an online dictionary should ideally be both **adaptive** (automatically adjusting to user input) and **adaptable** (open to the user's own personalization).

The Vocabulary.com Dictionary is fully integrated in an adaptive gameplay environment that quickly adjusts to the level of the user by means of *item response theory* (IRT), so that the learner's time is not wasted by vocabulary items too far above or below his or her proficiency. IRT is the central model in computerized adaptive testing (currently used in the US for such standardized tests as the GRE and GMAT), and it allows appropriate questions to be presented to test-takers, zeroing in quickly to their skill level based on their successive responses. For such an approach to work effectively, a large pool of questions must be continually calibrated, scored according to difficulty and discrimination (i.e., how well a question discriminates between users of different ability levels). Vocabulary.com has an ever-increasing pool of questions numbering more than 100,000, and the voluminous user data collected over millions of sessions provides accurate models for the appropriate questions to present to a vocabulary learner at any given time.

For each question, a graph is generated to plot the distribution of right and wrong answers across different ability levels, and a curve is drawn to fit the data. A curve that rises very

steeply indicates that a question is a good discriminator, accurately separating higher levels from lower levels of user proficiency. Below is an example of a highly discriminating question for the word *exuberant*. (Users' ability levels are plotted on the x-axis and the frequency of correct answers on the y-axis.)

The serenity of the fifth movement Andante is in stark contrast to the whirlwind of the _____ finale.

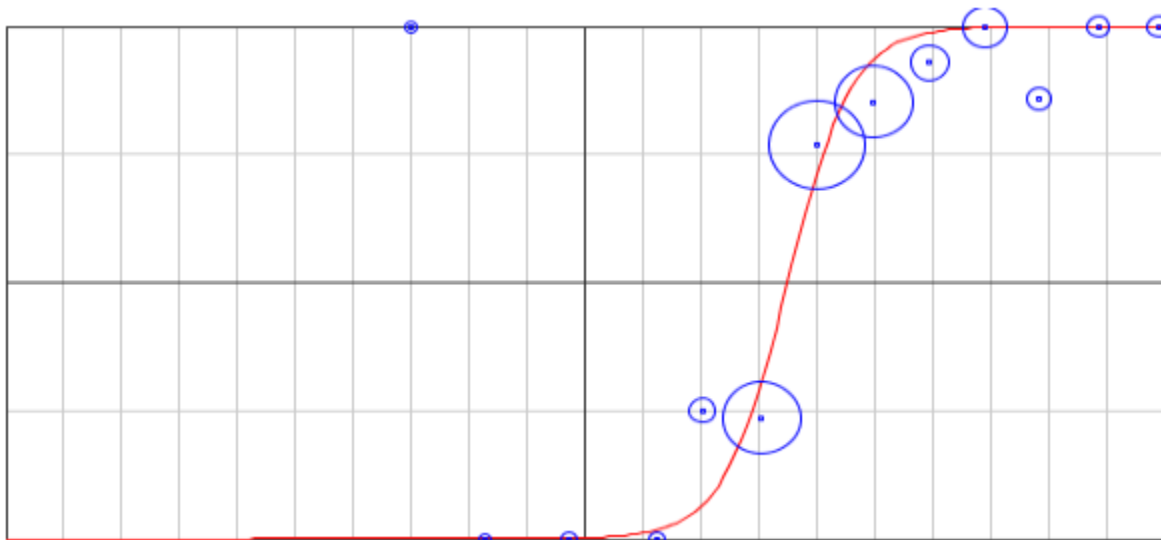
Source: New York Times

exuberant

equitable

indisputable

dormant



Such question graphs are continually redrawn to incorporate new player data, so that each question can be dynamically re-calibrated to assess its difficulty and discrimination as it is answered by more and more users.

The Vocabulary.com Dictionary is not only integrated into an *adaptive* learning system but is *adaptable* as well. Users can create their own custom-made vocabulary lists to study from and then learn those lists in the same gamified environment as the main Vocabulary.com Challenge. This also allows teachers to focus on those vocabulary items

that are appropriate to their students' particular needs. A learning environment that is both adaptable and adaptive can lend itself to vocabulary instruction in a way that transcends static flash cards and the rote memorization of word lists.

Harnessing the Power of the Corpus

The advent of *corpus integration* in lexicography has been transformational for creators of dictionaries and thesauruses, allowing real-world language data to paint pictures of words in naturally occurring contexts (Zimmer 2012). The creation and easy manipulation of corpora consisting of billions of tokens (individual instances of lexical items) opens up possibilities for “back-end” improvements in the building of dictionary entries, and also for innovative “front-end” features that allow users to plumb corpus data on their own. Rather than relying on the manual accumulation of citation slips, lexicographers can now consult a seemingly limitless supply of relevant text to illuminate the usage of any given item.

This embarrassment of riches presents new challenges, of course. Corpus data is most useful when lexical items are identified by part of speech; fortunately, automated part-of-speech taggers are making great strides in accuracy. But this type of pre-processing must still be accompanied by tasks that, at least thus far, only humans can adequately perform, such as “word sense disambiguation” (WSD) to determine which definition of a word in a dictionary entry corresponds with a usage in a particular sentence culled from the corpus.

The payoff from such semi-automated corpus work is sizable, however. Entries in the Vocabulary.com Dictionary are enriched by numerous corpus-derived example sentences that identify how words are used “in the wild,” as they appear in sources ranging from classic literature to contemporary journalism. Vocabulary.com’s corpus is constantly replenished by news feeds that can be scoured for appropriate example sentences using algorithmic methods. Here we see usage examples for *redolent* as they appear on the word’s page in the Vocabulary.com Dictionary:

USAGE EXAMPLES

[All](#) [News](#) [Medicine/Science](#) [Business](#) [Arts/Culture](#) [Sports](#) [Fiction](#) [Tech](#)

There are craft beers on tap and a Manhattan cocktail **redolent** of clove and orange peel.

Seattle Times May 22, 2013

His game, **redolent** of crisp iron play and clutch putting, has resurfaced.

New York Times May 4, 2013

Here, he visits Thailand to see a "trash fish" processing plant in action, with scenes **redolent** of Mr Burns's Animal Slurry manufacturing concern.

The Guardian Feb 28, 2013

Later I found Rocky sitting heavily on his bunk in a room that was **redolent** of graduate students with limited access to shower facilities.

Salon Dec 30, 2012

[Next](#) ►

Beyond this automatic generation of fresh corpus data for words in the dictionary, Vocabulary.com has enlisted trained lexicographers in a major WSD project to link corpus example sentences to appropriate meanings of a word. This allows Vocabulary.com to make use of such sentences as the basis for adaptive vocabulary instruction.

Vocabulary.com integrates corpus-derived sentences into vocabulary learning by generating a variety of question types. After the illustrated word is matched to the appropriate dictionary meaning, multiple-choice questions can be created that rely on the word's specific definitions and synonyms. In addition, sentences that exhibit sufficient contextual cues indicating a word's meaning can serve as the basis for fill-in-the-blank or choose-the-definition questions. Here is an example of such a question, in which the user chooses the best definition for the word that fits the context of the provided sentence example:

But on Friday afternoon, the first signs of open opposition appeared, with several leading political parties issuing strong **condemnations**. (Source: New York Times)

In this sentence, **condemnations** means :

- expressions of strong disapproval
- messages expressing an opinion based on incomplete evidence
- statements produced before a court of law
- accommodations in which both sides make concessions

A Dictionary for the Digital Native

In the digital age, creators of dictionaries need to adapt to their new online homes and understand the opportunities that web-based interaction presents. It is also vital that online dictionaries be enriched with data-driven and visually stimulating features in order to engage with “digital natives.” This means meeting the digital natives on their own technologized terrain, fulfilling expectations of interactivity and fluidity in sophisticated graphic interfaces.

Vocabulary.com demonstrates how an online dictionary can thrive, thanks to a combination of cutting-edge lexicography, computational power and web development. In reimagining the dictionary, Vocabulary.com honors the legacy of the print tradition while taking full advantage of the latest insights into online content and design. The content can now more accurately reflect language as it is actually used, with insights from corpus-driven methods illuminating semantic nuances as never before. Improvements in design can present this linguistic data more engagingly. And in so doing, Vocabulary.com is introducing new generations to the interwoven splendor of language.

REFERENCES

Atkins, Beryl T. Sue. 1996. “Bilingual Dictionaries: Past, Present and Future.” In: *Euralex’96 Proceedings*, Martin Gellerstam et al., ed. Gothenburg: Gothenburg University. 515-590.

Atkins, Beryl T. Sue. and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford University Press.

Dziemianko, Anna. 2012. “On the Use(fulness) of Paper and Electronic Dictionaries.” In: *Electronic Lexicography*, Sylviane Granger and Magali Paquot, ed. Oxford University Press.

Granger, Sylviane. 2012. “Introduction: Electronic Lexicography – from Challenge to Opportunity.” In: *Electronic Lexicography*, Sylviane Granger and Magali Paquot, ed. Oxford University Press.

Lannoy, Vincent. 2010. “Free Online Dictionaries: Why and How?” In: *ELexicography in the 21st Century*, Sylviane Granger and Magali Paquot, ed. Presses Univ. de Louvain.

Lew, Robert. 2012. “How Can We Make Electronic Dictionaries More Effective?” In: *Electronic Lexicography*, Sylviane Granger and Magali Paquot, ed. Oxford University Press.

McKeown, Margaret G. 1991. "Learning Word Meanings from Definitions: Problems and Potential." In: Paula J. Schwanenflugel, ed., *The Psychology of Word Meanings*. Lawrence Erlbaum Associates, 137-156.

Rundell, Michael. 2012. "The Road to Automated Lexicography: An Editor's Viewpoint." In: *Electronic Lexicography*, Sylviane Granger and Magali Paquot, ed. Oxford University Press.

Siegel, Muffy E.A. 2007. "What Do You Do with a Dictionary? A Study of Undergraduate Dictionary Use." In: *Dictionaries: Journal of the Dictionary Society of North America*. 23-47.

Zimmer, Ben. 2012. "Reconsideration: The Thesaurus." *Lapham's Quarterly* 5.2, Spring.
<http://www.laphamsquarterly.org/reconsiderations/word-for-word.php>