KWARTALNIK NEOFILOLOGICZNY, LVIII, 4/2011

MAŁGORZATA KOWALCZYK (GDAŃSK)

THE APPLICATIONS OF NEWS AGGREGATORS IN DICTIONARY MAKING

The aim of this paper is to present some of the practical applications of news aggregators in the process of dictionary making and to demonstrate that they can be useful tools in lexicographical practice. An *aggregator* is "a website that collects related items of content and displays them or links to them" (Stevenson 2010: 31). The paper is based entirely on Google News, one of the largest and freely accessible websites of this kind. While every aggregator is unique and different, they all share genetic similarities which allow generalizations concerning their applicability. Consequently, the applications listed below may be successfully implemented with regard to other similar websites. The presentation will involve the most important elements usually considered in dictionary making (Atkins & Rundell 2008: vi-vii, Svensen 2009: v-viii), which I also regularly come across in my lexicographic practice, such as: inclusion and validation, canonical form, grammatical identification, sense division, usage labels, usage examples, and word collocations – all amply illustrated using examples from Google News.

INTRODUCTION

News aggregators are vast archives of media texts from diverse sources which are in electronic form and made available online. Quintessentially, their sources are newspaper and magazine archives, although blogs and other unpublished speech material is also strongly featured. In this case, availability is limited to and synonymous with searchability; however, much depends on the search engine these aggregators are linked with, and much can be elicited and drawn from even simple searches. The most notable and most widely used news aggregator is Google News. In itself a part of a larger online resource platform, this is a computer-generated database holding a vast number of searchable media texts. It is composed principally of two types of resources. The first one is a news site which aggregates the latest media headlines from a vast number of English-language news sources worldwide, groups similar stories together, and displays them according to each reader's personalized interests. The second one is a highly extensive media archive feature, which offers access to historical archives going back at least several decades; there is also a timeline view, to browse news from various years. Although it is not a corpus developed to aid dictionary makers, it does have a number interesting advantages for lexicographers.

Aside from its no-cost accessibility, the biggest advantage of this aggregator is its powerful and sophisticated search engine employing the famous Google search algorthythms, which enables to make a variety of complex searches; these can be done according to words, phrases, contexts, wildcards, dates, times, as well as the type and provenience of sources. Another important advantage is the number and variety of texts stored in this aggregator: it draws from more than 25,000 English-language sources, which are enormously diversified in terms of subject (virtually all subjects are covered, from politics and sports to music and popular culture) and geographic origin of sources (virtually all countries where English-language media exist are represented, from the United States and Great Britain to India and Australia); because it is a repository of such a vast and diversified body of texts, it is an extremely powerful tool for linguists and lexicographers. Yet another advantage is that the holdings of its archives are constantly updated, even faster than in case of large corpora developed by consortia of universities and commercial publishers. Moreover, while the aggregator is constantly expanded to encompass the most recent data, care has also been taken to adequately store the old data and the archive goes back more than 200 years; this is invaluable in case of data from pre-computer era which has either been scanned or keyed into the computer; the aforementioned timeline view is a feature which enables browsing data from various years. All this has enormous implications for linguists and lexicographers.

INCLUSION AND VALIDATION

One of the most fundamental decisions the lexicographer has to make when starting the work on a dictionary is deciding which words to include and which to exclude from the dictionary. This is not always simple. Landau (2001: 202) says that some 800 or so new words come into common or working vocabulary of English every year; at the same time, numerous words fade in use and fall into oblivion. Aggregators come extremely useful in providing evidence that certain words exist in sufficient numbers to warrant inclusion in a dictionary. Put differently, the number of citations from diverse sources over a certain period of time helps the lexicographer determine the possible inclusion or exclusion of a word.

One of the prime considerations for inclusion and validation of words in the dictionary is determining their frequency of occurrence. News aggregators again be very helpful in assessing the frequency. In the past, frequency of words was often a matter of guesswork. Although frequency counts varies depending on the type of search engine, present-day aggregators offer much more precision and are simply incomparable with guesswork from the past. Even a cursory look at the very number of citations is revealing and helps determine whether or not to include a given word or sense in the dictionary. Below is a selection of randomly chosen words and the number of their occurrence in the Google News aggregator as of March 28, 2009, 17:39 (naturally, these are rounded-off figures):

the *300,000,000*	***food*** *11,900,000*	***eat*** *3,160,000*
of *274,000,000*	***sometimes*** *10,300,000*	***vacation*** *1,860,000*
is *187,000,000*	***English*** *7,290,000*	***intelligent*** *1,270,000*
it *155,000,000*	***sport*** *7,120,000*	***pizza*** *876,000*
not *117,000,000*	***crisis*** *5,670,000*	***delicious*** *729,000*
some *66,500,000*	***movie*** *4,600,000*	***cardiovascular*** *416,000*
what *59,600,000*	***baby*** *4,130,000*	***dude*** *222,000*
go *40,000,000*	***dance*** *3,390,000*	***shit*** *117,000*
war *42,900,000*	***sex*** *3,440,000*	***adjudicate*** *102,000*
money *23,800,000*	***cool*** *3,350,000*	***synthesizer*** *43,800*

Validation that a given word exists is extremely important in case of neologisms and ghost words. The decision whether to include them or not is possibly one of the most difficult. Again, aggregators come handy and offer solutions even with regard to expressions perceived as contrived or strange. Take, for instance, the word *walkperson* which, since the times of political correctness, has been occasionally used for a "Walkman personal stereo". The analysis of aggregatos results reveals that it this word actually lives its own life and by no means should be excluded from a dictionary:

*… Don't disturb my music on my **walkperson**. Hey, let me … (Boston Globe, 1991)*
*… by suggesting a cheap **walkperson** for the neighbor … (Chicago Sun-Times, 1993)*
*… What was she listening to on her Sony **Walkperson**? … (Literary Review, 1998)*
*… What's wrong with a Discman or a **Walkperson**? … (Guardian, 2008)*
*… political correctness front, as '**Walkpersons**' … (New York Times, 1992)*
*… I had my **walkperson** on, so I couldn't hear you … (Los Angeles Times, 1992)*
*… after jogging too long with my **Walkperson** on … (Boston Globe, 1993)*
*… by a furious female colleague to call it a **walkperson** … (Mirror, 2000)*
*… I listened to it obsessively on my **walkperson** … (New York Times, 2009)*
*… people took out their books or their **Walkpersons** … (Harvard Crimson, 1986)*

Let us take another example of an expression which has enjoyed a huge popularity, if only among the computerate: *ego-surfing* meaning "searching the Internet in hope of finding mentions of one's own name." Was this popularity short-lived? Clearly not, judging from these citations and their sources:

*… **Ego-surfing** is typing your name into a search engine and … (CBS News, 2002)*
*… I did some **ego-surfing** the other day and there was … (Rocky Mountain News, 2002)*
*… practice commonly known as **ego-surfing** or Googling yourself … (Wired News, 2005)*
*… is even a term, **ego-surfing**, to describe the phenomenon … (Chicago Tribune, 2006)*
*… **Ego-surfing** is okay once in a while, but I must lead a real sad … (Rediff, 2003)*

*… After a little **ego-surfing**, I found my column had been … (National Post, 2007)*
*… the obvious application is **ego-surfing**: watching reactions … (Info World, 2005)*
*… **Ego-surfing** is putting your name into a search … (Europe Intelligence Wire, 2003)*
*… it's commonly known as **ego-surfing** or Googling yourself … (Tampa Tribune, 2006)*
*… **Ego-surfing** can make you feel as small and insignificant as … (Aspen Times, 2008)*

Contextual citations also serve the purpose of illuminating the meaning of words. Take, for instance, the word *staycation* which simply means "a vacation spend at home or near home, especially because of insufficient funds to travel." Again, this neologism has enjoyed a huge popularity, most likely caused by the recent financial crisis. The citational corroboration provides us with context needed to understanding its meaning:

*… planning a **staycation** as opposed to a foreign holiday … (What's On Stage, 2009)*
*… the **staycation**, a vacation you take in your own home town … (ABC News, 2009)*
*… usually travels to France, is embarking on a **staycation** … (Newsweek, 2009)*
*… **staycation**, spending leisure time at home or close to home … (Omaha Herald, 2009)*
*… For Floridians, the concept of **staycation** is second nature … (Vero Journal, 2009)*
*… to make the most of your Bay Area **staycation** … (San Francisco Chronicle, 2009)*
*… A **staycation** is a getaway that is close to home. It's a great … (KLEW-TV, 2009)*
*… The family-themed **staycation** includes tickets and … (Des Moines Register, 2009)*
*… who are choosing a **staycation** by vacationing at home … (CBS News, 2009)*
*… to plan a budget-conscious Colorado **staycation** … (Denver Post, 2009)*

But validation is by no means restricted to new words; it can also refer to new meaning of the already existing words. Through the application of aggregators lexicographers may grasp the new sense of such words, identify them, and include in their dictionary. Such was the case with *grass* which has other meanings, including "marijuana":

*… workers who mow the **grass** at city parks … (Toledo On The Move, 2009)*
*… ganja shop and a lounge where people can smoke **grass** … (Seattle Post, 2009)*
*… seen in many neighboorhoods, overgrown **grass** and … (WDIV-TV News, 2009)*
*… **Grass** is regular at our parties, we don't consider it as a drug … (Times, 2009)*
*… arrested for smoking **grass**. I'm serious … (Washington Post Blogs, 2009)*
*… Fake **grass** has its place – the miniature golf course … (North County Times, 2009)*
*… The **grass** is always greener. Ain't that the truth? … (Brooklyn Papers, 2009)*
*… surfer dudes high on **grass** do not make the most stimulating … (Independent, 2007)*
*… lawnmowers and help tackle overgrown **grass** at several … (WTOL-TV News, 2009)*
*… smell of freshly cut **grass** permeates the suburbs … (Wall Street Journal, 2009)*

Another example of sense validation may be the word *kosher* which in informal English has a wider meaning of "acceptable or legal." Consider the following citations:

*… because they're getting a warm, **kosher** meal … (North Jersey Jewish News, 2009)*
*… Is it **kosher** to write someone's eulogy before they have died? … (Mom Logic, 2009)*
*… the nation's largest producer of **kosher** meat … (Jewish Telegraphic Agency, 2009)*
*… the package includes a **kosher** breakfast and … (San Francisco Chronicle, 2003)*

… if requested, ***kosher*** *food will be provided … (Baltimore Jewish Times, 2009)*
… ***Kosher*** *meat has enjoyed a reputation for high quality … (New York Times, 2009)*
… Is it ***kosher*** *to tell the truth about someone even if … (Oakland Tribune, 2006)*
… A ***kosher*** *breakfast is available for clients … (Jerusalem Post, 2009)*
… I'm still not sure whether it's ***kosher*** *to download and … (Info World, 2007)*
… It's ***kosher*** *to charge $8 for popcorn at movie theaters … (Denver Post, 2004)*

CANONICAL FORM

One of the most fundamental decisions the lexicographer has to make when starting to compile the list of entries is deciding upon the canonical form of words. The canonical form, let us recall, is "the form chosen to represent the headword of an article in a dictionary" (McArthur 1992: 188). It has to do either with the most representative spelling or the most representative grammatical form of the word. News aggregators come very handy when it comes to determining the canonical form. Let us make a simple comparison of a dozen or so words which have two alternate forms and see how these forms translate statistically into their occurences in the Google News aggregator (the comparison was done on March 28, 2009, at 18:49):

adapter *261,000*	***adaptor*** *26,700*
complextion *193,000*	***complection*** *246*
doggie *42,700*	***doggy*** *35,600*
enology *4,270*	***oenology*** *1,240*
fetus *119,000*	***foetus*** *22,000*
gauge *712,000*	***gage*** *306,000*
jail *3,620,000*	***gaol*** *66,400*
orthopaedic *61,300*	***orthopedic*** *201,000*
realization *413,000*	***realisation*** *325,000*
sulphur *182,000*	***sulfur*** *143,000*

As one can see, even a cursory quantitative analysis of the results may suggest potentially which of those forms to chose as canonical. A similar comparison can be done with regard to nouns which have two plural forms:

agendas *238,000*	***agendae*** *63*
appendices *40,800*	***appendixes*** *9,110*
cactuses *6,190*	***cactii*** *134*
condominiums *270,000*	***condominia*** *109*
goyim *2,230*	***goys*** *919*
aparazzi *120,000*	***paparazzos*** *166*
radii *13,200*	***radiuses*** *779*
referendums *87,400*	***referenda*** *36,600*
tableaux *88,100*	***tableaus*** *7,680*
tuxedos *34,100*	***tudexoes*** *3,120*

News aggregators can also be helpful in eliciting information about most frequent spellings in a given variety of English, and are thus instrumental in establishing the canonical form in these varieties. Take for instance the word *color*, which forms suggests American spelling; if one enters *color* in search engine, one is likely to get the following results, all of which come from American sources:

*… or dark brown in **color** and are easier to spot. Bed bugs are … (Examiner, 2009)*
*… in order not to compromise the **color** of the paintings … (USA Today, 2009)*
*… glowed with their unique reddish-orange **color** … (Ventura County Star, 2009)*
*… **color** copies ar Kinko's start at 20 cents a sheet … (Wichita Eagle, 2009)*
*… it's not gray in **color**, it's something else … (San Jose Mercury News, 2009)*
*… which gives accurate, consistent and predictable **color** … (Quick Printing, 2009)*
*… solution is to choose a dark **color** … (Atlanta Journal Constitution, 2009)*
*… image works better in **color** than in black and white … (Kansas Online, 2009)*
*… she didn't like the **color**, which at the time was … (Houston Chronicle, 2009)*
*… right to decide which **color** is best? Assigning one … (Hartford Courant, 2009)*

By analogy, a search for an alternative spelling of the same word, that is, *colour*, might yield the following results, all of which come from British sources and help us decide on the right canonical form:

*… because it's in pale blue, a beautiful **colour** which … (Oxford Mail, 2009)*
*… there is something about the use of **colour** and the impact … (Independent, 2009)*
*… in allowing a man of **colour** to be the absolute best … (Daily Monitor, 2009)*
*… if you wefre working with light **colour**, you would use … (BBC News, 2009)*
*… dark **colour** shows off your items much more beautifully … (Telegraph, 2009)*
*… sticky or dirty and gives a nice natural **colour.** In my own … (Mirror, 2009)*
*… Contrast and **colour** reproduction are also good … (PC World Magazine, 2009)*
*… goes on a silver blue **colour** which lights up your … (Glasgow Daily Record, 2009)*
*… with no **colour**, perfume or harsh chemicals they are ideal for … (Mirror, 2009)*
*… with the skin the **colour** of alabaster, I shan't be removing … (Telegraph, 2009)*

Deciding on the canonical form may be especially important in case of phrases and longer expressions, which expecially in American lexicographic practice are not nested but located in the dictionary under the first letter of the first word (Landau 2001: 107). For instance, the expression *cold turkey,* referring to "a sudden discontinuation, especially of a habit," may be used with a number of verbs such as *go, quit, stop, drop,* etc; the analysis of results will help establish which verb is most frequent, and thus, what the the canonical form of the entire expression should look like; judging from the citations below this verb seems to be *go*:

*… he used steroids and **went cold turkey** and needed … (Associated Press, 2009)*
*… Don't **go cold turkey**. Abrupt caffeine withdrawal can … (Los Angeles Times, 2009)*
*… came up with a brilliant solution: have him **quit cold turkey** … (Examiner, 2009)*
*… if the government **dropped** it **cold turkey**, it could … (Moonee Valley News, 2009)*
*… drug addict can painlessly **go cold turkey,** overcome … (Chattanooga Times, 2009)*
*… she'd checked herself into a facility and **gone cold turkey** … (Deseret News, 2009)*
*… You should **go cold turkey** on self-help articles … (US News & World Report, 2009)*

*… **I did** it **cold turkey**, says Miller, who lives in … (Lower Hudson Journal, 2009)*
*… of Internet addictions but I found **going cold turkey** on … (Seattle Times, 2009)*
*… addict trying **stop cold turkey**. Suleman daid doctors … (National Post, 2009)*

Similarly, the expression *call it a day* is sometimes heard to be used with other words which replace the element *night* to form such expressions like *call it a night* or *call it an evening* The analysis of results will help us determine what these other words are and, statistically, if such combinations are mere variants or if they perhaps should be treated as a canonical form (which, judging from the citations below, is not the case):

*… they should do a wrap-up and **call it a day** … (Entertainment Weekly, 2009)*
*… the couple decide to **call it a night** and head for home … (Yuma Sun, 2009)*
*… we would **call it a day** but there's plenty of reason to believe … (Time, 2009)*
*… by the time we decide to **call it an evening**, the heat … (Charleston Post, 2009)*
*… and **call it a day**, but I'd like my kids's summers to be … (Washington Post, 2009)*
*… staff persuaded him to **call it a day**. Those who … (Spokesman Review, 2009)*
*… Should the band just **call it a day**? No! Not Queen! … (Entertainment Weekly, 2009)*
*… said he was just ready to **call it an evening** … (Dallas Morning News, 2004)*
*… we did that, so let's **call it a day** … (San Antonio Express-News, 2002)*
*… we're out of pitchers, so let's **call it a day** … (San Francisco Chronicle, 2002)*

Sometimes the analysis of results is surprising and shakes our beliefs about what the canonical form should be. For instance, the colloquial expression *not to give a damn* meaning "to be indifferent to or contemptuous of" is usually entered in dictionaries exactly in this form, that is, in the negative. However, evidence suggests that it not always used in negation, and that it is often used in the affirmative. Here is the citational corroboration which contains both negative and positive occurences of the said phrase:

*… they're lazy, don't **give a damn**, don't pay attention … (Toronto Star, 2009)*
*… They do **give a damn**, but not the way most people … (Washington Post, 2009)*
*… We want these games, and we don't **give a damn** if … (Daily Northwestern, 2009)*
*… You actually **give a damn** about what happens to him … (Broward Times, 2000)*
*… He won't **give a damn** about its expensive features … (Business Week, 1995)*
*… that will never **give a damn** to anyone's suffering … (New York Times, 2009)*
*… Unlike Rhett Butler, he does **give a damn** … (Chicago Tribune, 1988)*
*… Frankly my dear, I don't **give a damn** … (New York Newsday, 2009)*
*… now everyone **gives a damn** about Africa … (Entertainment Weekly, 2009)*
*… got to the point where they don't **give a damn** … (Minnesota Independent, 2009)*

GRAMMATICAL IDENTIFICATION

Perhaps one of the less evident applications of aggregators in dictionary making is establishing of grammatical identification of words, which is nevertheless crucial in any dictionary. This application is visible in relation to a number of grammar-related issues.

One such issue is functional shift, alternatively referred to as zero derivation. Functional shift involves the change of grammatical category (part of speech) without the change in form. Due to its analytic nature, the English language is extraordinarily amenable to this process. Still, not every English word is likely to undergo it. Browsing through lexical evidence enables lexicographers to find instances of functional shift. Consider the noun and the verb *mister* in these citations:

… *Do I call him* ***mister*** *or by his first name? … (Bangor Daily News, 2006)*
… *hey,* ***mister****, you're going the wrong way … (St. Louis Post-Dispatch, 2009)*
… *Why should I* ***mister*** *him? He's no better than me … (Google Books, 2009)*
… *don't* ***mister*** *them in this office. Do you understand? … (Google Books, 2009)*
… *They address you as* ***mister*** *this,* ***mister*** *that … (Seattle Times, 2009)*
… *and didn't everyone call him* ***mister****? Let's see … (Washington Post, 2003)*
… *anymore. Listen,* ***mister****, there's no time to … (Chicago Tribune, 2000)*
… *Forget the lights,* ***mister****, this is an emergency … (Sunday Herald, 2009)*
… *don't* ***mister*** *me, lady. Just call me Mac … (Google Books, 2009)*
… *by the way, many called him* ***mister*** *… (Charleston Post Courier, 2009)*

Another example of functional shift, from verb to noun, involves the word *go* which changed grammatical category and also changed its meaning to "activity or operation" or "try or endeavor" or "energy." See the following citations:

… *He had refused to* ***go*** *home for thirty-two years … (New Vision, 2008)*
… *If you're an American and you* ***go*** *to Europe in the winter … (Jaunted, 2009)*
… *this will be a van with plenty of* ***go*** *and stuff … (Australia News, 2009)*
… *Jordan left basketball to take a* ***go*** *at baseball … (Chicago Tribune, 1996)*
… *For more information on Jackson,* ***go*** *to his Web site … (Tampa Tribune, 2009)*
… *Slocomb residents wondered if the project was still a* ***go*** *… (WSFA News, 2009)*
… *it will encourage more people to* ***go*** *around the Cape … (ABC News, 2009)*
… *The July 30 performance in Oslo is a* ***go****. Tour producer … (Ticket News, 2009)*
… *Sometimes you can* ***go*** *over the line. I don't know if … (USA Today, 2009)*
… *I don't think he should* ***go*** *to America because there are … (BBC News, 2007)*

Functional shift is much less frequent in other parts of speech. Nevertheless, by analyzing citations from news aggregators, lexicographers may find some interesting part-of-speech conversions. Consider the following examples of word *off* which changed its grammatical category from preposition to verb, and changed its meaning to "to kill":

… *the presence of the US warship apparently scared them* ***off*** *… (Guardian, 2009)*
… *saw two men fall* ***off*** *the cliff near Sunset Bvd … (San Diego Tribune, 2009)*
… *by making gay puns about Vito and vowing to* ***off*** *him … (Boston Globe, 2006)*
… *forced to clear* ***off*** *the line in a desperate second half … (Independent, 2009)*
… *The bombs didn't go* ***off****. Since 1999, many people have … (USA Today, 1999)*
… *He wants to* ***off*** *her for the insurance money … (New York Newsday, 2007)*
… *right to send them* ***off*** *without a proper name … (New York Daily News, 2009)*
… *The house is raised* ***off*** *the ground, facing … (San Jose Mercury News, 2009)*

… is to ***off*** *them yourself, making you a killer … (CNET News, 2007)*
… you can't turn it ***off*** *without rebooting the PC … (Atlantic Monthly, 2009)*

Sometimes an aggregator helps to determine grammatical behavior of words and check these against the linguistic reality. For instance, the noun *crazy* appears as a canonical form in most dictionaries. The analysis of aggregator concordances of this word reveals something striking:

… they should pull over and let the ***crazies*** *pass … (Bakersfield Californian, 2009)*
… I'm not one of those ***crazies*** *who thinks we ought to put … (Moderate Voice, 2009)*
… Gabriel, a ***crazy*** *who may just have more … (Los Angeles Times, 1990)*
… The ***crazies*** *now are not carrying guns but legislation … (Politics Daily, 2009)*
… second story was about a ***crazy*** *who also happened to … (Kansas City Star, 1992)*
… Even the talk show ***crazies*** *will have trouble generating … (Boston Globe, 2009)*
… there was a man, a ***crazy*** *who lived upstairs … (Los Angeles Times, 1991)*
… notice extremes: the ***crazies*** *on both ends … (Las Vegas Review-Journal, 2009)*
… you'd have some ***crazies*** *attacking their neighbors … (American Chronicle, 2009)*
… too many crooks and ***crazies*** *on the streets these days … (MSNBC-TV News, 2009)*

The overwhelming majority of all concordance hits for the noun *crazy* are in plural. Based on this evidence lexicographers may conclude that this noun is usually used in the plural and may wish to signal this in a dictionary.

SENSE DIVISION

Aggregators are also applicable in helping to determine whether to include an individual sense of a word in the dictionary and if so, whether to put it in the prominent position in the headword or place it at the end of the entry, or whether to enter it as one, or perhaps split into two or more sense. Much depends on the individual preferences of the lexicographer or publisher's guidlines regarding a particular dictionary. Of course in this case a simple count of occurrences will not suffice and one has to analyze each individual citation one by one. Still, the analysis of concordances may sometimes reveal additional, less obvious senses of the word and even a seemingly uniform meaning of a word may need to be appended with additional senses. A serviceable example is the slang word *nerd* which roughly means "a person lacking in social skills, fashion sense, or both":

… It is about a ***nerd*** *and a popular girl who fall in love … (Cinema Blend, 2009)*
… I'm a ***nerd****: I love technology. I'm surrounded … (Guardian, 2009)*
… She's dressed like a ***nerd****! She always dresses like that … (Manitoban, 2001)*
… locker room and told me I looked like a ***nerd*** *… (Big Ten Network, 2009)*
… ***nerd*** *lacking social skills, know-it-all … (America Intelligence Wire, 2006)*
… The shy ***nerd*** *slipped out of sight again … (St. Louis Post-Dispatch, 1992)*
… You don't have to be a ***nerd*** *to study science … (Popular Science, 2009)*

… *Clark Kent was a pretense, a shy and clumsy* ***nerd*** *who … (Miami Herald, 2002)*
… *or get your local computer* ***nerd*** *to configure it for you … (Stuff, 2009)*
… *He dresses like a* ***nerd*** *and talks like a* ***nerd*** *… (Telegraph, 2003)*

As evidenced above, there is another semantic dimension of this word, evident in the frequent association with two spheres: a "bookish person, usually hard studying before the exam" and "a person with fascination with or expertise in computers." This makes one re-consider the original definition and perhaps either broaden it or to add two additional senses encompassing the afore-mentioned two dimensions.

Similarly, let us take the word *ghetto*. It was first used in English to mean a "part of a European city where Jews were forced to live." The meaning, however, was later extended in the United States to a "rundown or overpopulated part of a city where a minority group, especially, African-Americans, lived." Still later, the word became to mean any place of metaphorical confinement.

… *she managed to escape from a* ***ghetto*** *where Jews were … (New Jersey Standard, 2009)*
… *the story begins in a* ***ghetto*** *in East Baltimore and ends … (Baltimore Sun, 2009)*
… *brutalized during her teenage years in the Warsaw* ***ghetto*** *… (Tampa Tribune, 2009)*
… *film got him a ticket out of the urban* ***ghetto*** *… (San Francisco Chronicle, 2003)*
… *Jews attempting to flee from the Vilnius* ***ghetto****. They were … (Reuters, 2009)*
… *dealers and armed gangs of a Los Angeles* ***ghetto*** *and turned them … (People, 2003)*
… *relocated to a Black* ***ghetto*** *which was West Oakland … (Examiner, 2009)*
… *it is a huge upper-middle-class* ***ghetto****. Ironically, the parents … (Regen, 2009)*
… *on the right is Jehuda Widawski,* ***ghetto*** *surviror from Israel … (Shtetlinks, 2009)*
… *base beyond the white middle-class* ***ghetto*** *that has … (Los Angeles Weekly, 2003)*

News aggregators are also applicable in determining the exact meaning because of the citational context. For instance the word *baddie,* which basically means a "bad person," is usually applied to film characters. The contextual citations prove it:

… *goes on to name-drop such* ***baddie****s as Charles Manson … (Pop Matters, 2001)*
… *Bana jumped at the chance to be a* ***baddie*** *in Star Trek … (STV-TV News, 2009)*
… *watching Dean and thinking: I wanna be a* ***baddie*** *and when … (Northern Echo, 2009)*
… *he is a* ***baddie*** *from the serial-killer series 'Dexter' … (Chicago Sun-Times, 2009)*
… *and slash a* ***baddie*** *halfway across the screen … (PC World Magazine, 2009)*
… *will have lasers and a* ***baddie*** *who shoots at the good guys … (Torontoist, 2009)*
… *Grant is up to play a* ***baddie*** *in the episode … (Hollywood News, 2009)*
… *and they featured promimently as* ***baddie****s in Bond films … (Forbes, 2009)*
… *I find it such a shame that they are such* ***baddie****s. They have … (Telegraph, 2009)*
… *Humans are resource-hungry* ***baddie****s. Period. But this … (Denver Post, 2009)*

Sometimes new senses can be discerned from citational evidence. For instance, the word *situation* in the American English has acquired another sense "problematic or troublesome situation." Again, this is corroborated by the contextual citations:

… work in the league, in each ***situation*** *he's been in … (Washington Post, 2009)*
… We have a ***situation*** *in this country with respect to violence … (OpEd, 2009)*
… The Chrysler ***situation*** *is somewhat different … (Detroit Free Press, 2009)*
… It's a desirable ***situation*** *to be in leading up … (New York Newsday, 2009)*
… with the team sinking and its financial ***situation*** *shaky … (USA Today, 2009)*
… We have a ***situation****. Can you come and help? … (San Jose Mercury News, 2009)*
… The Red Cross continues to monitor the ***situation*** *… (Easton Courier, 2009)*
… you just don't want to inflame the ***situation*** *… (Chicago Tribune, 2009)*
… I have a ***situation*** *here and I need your help … (Las Vegas Sun, 2003)*
… I have a ***situation*** *on the third floor … (Worcester Telegram Gazette, 2005)*

Accordingly, the lexicographer should consider adding another sense of the word *situation* and append it with the label "American" or "chiefly American."

Connotation is especially important to the foreign learners of English. News aggregators can be again extremely helpful in elucidating the connotative nuances. Take, for instance, the word *collossal* and consider its immediate lexical surrounding:

… Mount St. Helen's ***colossal*** *eruption on May 18 sent a huge … (Reuters, 2009)*
… he displayed a ***colossal*** *intellectual ignorance … (Politico, 2009)*
… representing a potentiall ***colossal*** *loss for American … (Washington Post, 2009)*
… The Treasury Department has a ***colossal*** *task to … (Washington Independent, 2009)*
… or another leading to a ***colossal*** *collapse or dramatic … (Kansas City Star, 2009)*
… Deregulating financial services was a ***colossal*** *mistake … (Detroit News, 2009)*
… it was also a ***colossal*** *waste of money. In these times … (New York Times,, 2009)*
… Thanks to Blago's ***colossal*** *screw-ups, he's now the … (Time Out Chicago, 2009)*
… he turned marble and stone into ***colossal*** *monuments that … (Miami Herald, 2001)*
… combination of the companies is be a ***colossal*** *undertaking … (Business Week, 2009)*

When one scans the list of citations contaning this word, it is apparent that there are numerous negative contexts of the figurative examples. In any EFL/ESL dictionary this is extremely valuable information.

USAGE LABELS

Aggregators provide useful information concerning labeling of words in dictionaries. Usage labeling gives information about the stylistic use of words, and labels include such styles as formal, informal or colloquial, slang, vulgar, sub-standard, non-standard, dialectal, archaic, outdated, and old-fashioned. It may also provide information about the scope of usage with regard to American vs. British English. For instance, the word *Slavonic* tends to be used in British English and not in American English; citational evidence supports this assumption and a random hit reveals no American sources with this word:

*… cultural potential of the **Slavonic** languages is huge … (Independent, 2004)*
*… were brought up in the spirit of **Slavonic** culture … (History Today, 1997)*
*… All the **Slavonic** countries have Russian as shared language … (Guardian, 2002)*
*… degree at School of East European and **Slavonic** Studies … (Independent, 2006)*
*… course at the department of **Slavonic** studies at Cambridge … (Guardian, 2006)*
*… treated as their counterparts from **Slavonic** countries … (BBC News, 2000)*
*… the status of **Slavonic** studies rose as governments realised … (Times, 2004)*
*… an appeal in favour of two **Slavonic** languages being … (New Statesman, 2003)*
*… He'd introduce his guest to the pub with a **Slavonic** accent … (BBC News, 2000)*
*… Observers like Macaulay of the School of **Slavonic** Studies … (BBC News, 2000)*

By contrast, let us take the word *movie,* which is consistently labeled in dictionaries as American, but has crept into British English. This may be testified by the following citational examples, which feature British sources as well:

*… wrote the script for Nichols's **movie** 'Heartburn' … (New York Times, 2009)*
*… The **movie** is about dysfunctional relationships … (Chicago Sun-Times, 2009)*
*… The **movie** 'Rebel Without a Cause' was originally meant to … (Times, 2009)*
*… to guarantee that a **movie** will be done on time … (Los Angeles Times, 2009)*
*… The **movie's** plot is a combo platter built upon two … (USA Today, 2009)*
*… the UK Film Council said: 'It's about quality **movies'** … (Independent, 2009)*
*… of the vampire love **movie** 'Twilight' … (Seattle Post-Intelligencer, 2009)*
*… **movie** 'Let The Right One In', released in UK cinemas … (BBC News, 2009)*
*… as the **movie** that won Winslet a well-deserved … (Washington Post, 2009)*
*… the $2-million **movie** about actors auditioning … (Los Angeles Times, 2009)*

Another example is the exotic-sounding word *naartjie*, which is South African English for a "tangerine". Note that the majority of sources are, of course, South African:

*… I hired someone to pick the **naartjie**s and sell them … (South African Times, 2007)*
*… flavored with **naartjie**s, native tangerines … (Atlanta Journal-Constitution, 2000)*
*… they could do with oranges, **naartjie**s and apples … (South African Star, 2003)*
*… and eating bananas and **naartjie**s. We all … (Sunday Independent, 2006)*
*… sells oranges and **naartjie**s to more than 50 markets … (Business Report, 2006)*
*… it explains why so few **naartjie**s are dispatched onto the … (Cape Times, 2006)*
*… Place **naartjie**s in a large glass bowl and leave to … (South African Star, 2005)*
*… South African **naartjie** exports to the US are … (South Africa Online, 2004)*
*… and which are called **naartjies**. The fruit … (Globe and Mail, 2006)*
*… The dessert is a uniquely South African **naartjie** pudding … (Independent, 2008)*

Yet another example is the medical-sounding word *biodata*, which is Indian English (or, more broadly, Asian English) term for a "curriculum vitae":

*… such eligible graduates should send their **biodata** and … (Hindu, 2005)*
*… asked to ensure that **biodata** of the applicant was attached … (Expressindia, 2006)*
*… According to your **biodata**, you made your acting debut … (Philippine Star, 2009)*
*… Her **biodata** says she is Muslim. To say one is … (New Straits Times, 2005)*
*… Applications containing **biodata** should be sent before … (Ceylon Daily News, 2009)*
*… Starting with my **biodata**, to recession and the role … (Kerala Online, 2009)*
*… I sent my **biodata** to Amitabh Bachchan stating my desire … (Expressindia, 2006)*

… person's ***biodata*** *is a valid way to look for good workers … (Economic Times, 2006)*
… According to Thakre's ***biodata****, around 3742 photographs … (Indian Express, 2007)*
… that her ***biodata*** *had been sent to the same bureau … (Times of India, 2007)*

Currency of a given word or its meaning is of a prime importance in dictionaries. News aggregators serve as tools to attest that a given word is actively used, or determing whether it is merely old-fashioned, or perhaps completely archaic. One can learn that, for instance, the most citations of the word *speakeasy* (meaning "a cheap or illegal saloon where alcohol is consumed") come from 1920s and 1930s and then, surprisingly, from the beginning of 20th century. The same goes with *bootlegger* meaning "a person who sell illegal whiskey" which has the highest number of citations from 1920s and 1930s. Let us have a look at the word *gook*, which does not refer only to the Vietnamese, but during various times in history, referred to other East Asians as well:

… McCain referred to his Vietnamese captors as ***gooks*** *… (Independent, 2007)*
… Chinese or Japanese, or anyone from Asia, were ***gooks*** *… (Atlanta Journal, 2007)*
… the Vietnamese as ***gooks*** *during the Vietnam War … (Los Angeles Times, 2007)*
… most Koreans take no offense when called ***gooks*** *by GI's … (Chicsgo Tribune, 1987)*
… for whom Chinese people are never Chinese, they're ***gooks*** *… (New York Times, 1992)*
… referred to Vietnamese as ***gooks****. For a presidential candidate … (Newsweek, 2008)*
… they called us ***gooks*** *and dirty Japs, but we knew … (Wadhington Post, 1995)*
… I believe that the Vietnamese also were called ***gooks*** *… (Washington Post, 1998)*
… anger erupted against Koreans and Chinese, called ***gooks*** *… (St. Paul Press, 2009)*
… officers and even refer to the Koreans as ***gooks*** *… (New York Times, 1990)*

USAGE EXAMPLES

Usage examples, that is illustrative quotations that exemplify the usage of particular senses, are a critical part of the dictionary entry and become even more important in case of EFL/ESL dictionaries. First of all, they convey a great deal of information about the grammatical context (such as transitivity of verbs, countability of nouns, or gradability of adjectives and adverbs), stylistic usage (such as the register or level of formality), connotation (such as affective implications), and naturally, referential or designative meaning. Often there is no better way to provide such information than by providing appropriate examples. The high number of citations used in this paper proves this point.

WORD COLLOCATIONS

Collocations are commonly co-occuring words. Put differently, they are words that usually accompany each other, possibly because of traditional, timehonored usage. Collocations are hardly visible in traditional citation files but can easily be discovered through aggregators. They are also of enormous importance to the foreign learners of English, and constitute a key part of any EFL/ESL dictionary. Take, for instance the word *abysmal*. What are its most common collocates? Even a cursory analysis of citations provides the answer to this question:

… demonstration of an ***abysmal lack*** *of understanding … (Charlottesville Daily, 2009)*
… a look at the ***abysmal conditions*** *that the students have to … (ABC-TV News, 2009)*
… Congress increase in the ***abysmal level*** *of foreign aid … (Time, 2002)*
… One more aspect of the ***abysmal stupidity*** *of Bush and company … (CBS News, 2008)*
… this was the ***abysmal situation*** *confronting … (New York Newsday, 2009)*
… by government forces and living in ***abysmal conditions*** *… (Voice of America, 2009)*
… the source of the ***abysmal situation*** *in Afghanistan … (World Meets, 2009)*
… fallen below the ***abysmal level*** *of their male peers … (New York Times, 2009)*
… the government to highlight its ***abysmal record*** *… (Wall Street Journal, 2009)*
… Only ***abysmal stupidity*** *could lead Meese to make any … (Los Angeles Times, 1987)*

Let us take another example. The slang verb *to bogart* used to mean “take more than one’s share or fail to share with, specifically with regard to a cigarette or marijuana cigarette”. The analysis of citations reveals that a number of collocates widened and this verb is used in a more general sense:

… he’s the one who’s been ***bogarting*** *all the tacos … (TV Online, 2006)*
… broke out over the cheerleaders ***bogarting*** *a bathroom … (Seattle Post, 2005)*
… his frustration at being unable to ***Bogart*** *a hotdog … (Jewish Week, 2009)*
… Times ***bogarted*** *all of the discussion this morning … (New York Times, 2008)*
… Johnson could ***bogart*** *a joint while modeling his new … (Albuquerque Journal, 2009)*
… genre where the white man has ***bogarted*** *the best roles … (Edmonton Sun, 2008)*
… Don’t ***bogart*** *that candy bar, my friend … (Palm Beach Post, 2007)*
… Don’t ***bogart*** *that joint was a way of telling … (Atlanta Journal, 1994)*
… ***bogarting*** *their machines during work hours, and … (Washington Post, 2005)*
… the device was being ***bogarted*** *by a team manager … (Register-Guard, 2007)*

Additionally, aggregators offer a possibility to analyze other combinations of words which border on collocations and lexicalized noun phrases. Take, for instance a colloquial word *bunny* meaning “a person, especially a young woman, who has a specified characteristics or nature to a high degree” or “a devotee or enthusiast, especially young woman, of a particular thing or of people who deal with that thing”:

… it’s too hot and I’m not a ***beach bunny****. I like … (Maui News, 2009)*
… She says: I won’t be a ***sex bunny*** *… (Wall Street Journal Online, 1999)*
… workers may have liked being called ***jungle bunny*** *… (Chicago Sun-Times, 2004)*
… he rides the waves, makes love with a ***snow bunny*** *… (Globe News Wire, 2009)*
… wrap up and take home as a ***cuddle bunny*** *… (Philadelphia Inquirer, 2006)*
… by the end I’m one tired little ***ski bunny*** *… (Philadelphia Weekly, 2006)*
… that her film debut marked as a ***sex bunny*** *… (Philadelphia Daily News, 1991)*
… to see commercials with an adorable ***snow bunny*** *… (Woodbridge Sentinel, 2009)*
… a fresh-faced ***sex bunny*** *shows up to seduce the new … (Entertainment Weekly, 2001)*
… an African-American female as a ***jungle bunny*** *in the presence … (New Jersey, 2004)*

REFERENCES

ATKINS, S. / RUNDELL, M. (2008) *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press

LANDAU, S. (2001) *Dictionaries: The Art. and Craft of Lexicography*. Cambridge: Cambridge University Press

MCARTHUR, T. (1992) *Oxford Companion to the English Language*. Oxford: Oxford University Press

STEVENSON, A. (2010) *Oxford Dictionary of English*. Oxford: Oxford University Press

SVENSEN, B. (2009) *A Handbook of Lexicography*. Cambridge: Cambridge University Press