The Complementarity of Crowdsourced Dictionaries and Professional Dictionaries viewed through the Filter of Neology

Franck Sajous, Amélie Josselin-Leray, Nabil Hathout

▶ To cite this version:

HAL Id: halshs-01955894
https://halshs.archives-ouvertes.fr/halshs-01955894v2

Submitted on 10 Jul 2019

The Complementarity of Crowdsourced Dictionaries and Professional Dictionaries viewed through the Filter of Neology

Franck Sajous, Amélie Josselin-Leray and Nabil Hathout
CLLE-ERSS, CNRS & Université de Toulouse 2

Abstract: This paper shows how amateur dictionaries, through their distinctive features (shorter inclusion timespan, analysis of the motivation, broad spectrum of expertise among the contributors) can complement professional dictionaries for the treatment of neologisms.

Keywords: semantic neology, lexical creations, lexicography, folk linguistics, motivation, subcultures, lexical competition, Urban Dictionary, Wiktionary

Introduction

This paper is part of a series of papers aiming to compare dictionaries compiled by professional lexicographers to dictionaries written by so-called amateurs. In previous studies, we have compared the macrostructures of those dictionaries and their coverage of the lexicon of various fields [Sajous et al., 2014], analyzed the idea of neutrality in relation to the informativeness of definitions [Sajous and Hathout, 2017] and compared their treatment of a specialized field – computing science [Sajous et al., 2018]. In those papers, we have shown the specific features of the two types of dictionaries and their complementarity instead of simply opposing them. In this paper, the comparison of professional and amateur dictionaries is viewed through the filter of neology. Such a perspective can seem paradoxical since one usually considers that once a lexical unit has been recorded in the dictionary, it loses its neological status [Sablayrolles, 2008; Mortureux, 2011]. In reality, things are slightly more complex. First, there is no real consensus on the neological status of lexical units. Second, when professional dictionaries, just like amateur ones, compete to include the latest buzzwords (which are considered a selling point), the semantic descriptions of the new headwords included in the macrostructure are not always accurate since lexicographers do not have the benefit of hindsight. Another thorny issue in the study of neology and, as a consequence, in the lexicographical treatment of neologisms, is the wide array of linguistic and extralinguistic phenomena inducing the change. Despite

the many studies on that topic, there is no universally acknowledged typology of neologisms. The words under study, through the comparison of a number of definitions taken from professional general-purpose dictionaries and from amateur dictionaries on the one hand, and from term banks and specialized amateur dictionaries on the other hand, have been chosen to illustrate both the various difficulties of dealing with neologisms in dictionaries and the wide range of linguistic phenomena that lead to linguistic changes. The words analyzed are either relatively recent neologisms (e.g. *fake news, post-truth*) or words that are not formal neologisms but which may fall into the following categories: (i) semantic neologisms (e.g. *hacker* in English), (ii) words that refer to evolving concepts (e.g. *hackathon* is not considered a neologism any more in the field of computing science – even if the definitions provided for this term in the field are not always accurate –, but its determinologization has brought about new meanings), (iii) words that are back in the front-page news and are still controversial and thus subject to metadiscursive statements, which in the end leads to new semantic shifts (e.g. *migrant / refugee* in British English), (iv) words that are specialized but whose technicity is rather low (e.g. *hackathon* again) or words that come from subcultures whose deep knowledge is a prerequisite for lexicographers trying to identify relevant semantic features (e.g. *graphic novel / roman graphique*), (v) words which are subject to lexical competition and whose very existence can only be accounted for by the analysis of their motivation and by identifying the linguistic actors at play (e.g. *pro-life* and *anti-abortion*), the issue being that the conclusions of such an analysis are often deemed too controversial to be explicitly stated in dictionaries. Although our main focus is on English neologisms, we also occasionally comment on French neologisms when they result from borrowing or are a literal translation that is used either at the same time as the English neologism or shortly later: in that case, analyzing whether the differences in the treatment of those neologisms are due to different lexicographical practices, to different sociocultural contexts or to a shift in the concept that the neologism relates to can prove enlightening.

Section 1 is dedicated to the presentation of the dictionaries under study, with a particular focus on crowdsourced ones, whose specificities (editorial policy, contribution mode, etc.) are relevant for the issues at stake in the study of neology. In section 2, we study some examples of euphemisms and oxymorons that were not coined for stylistic purposes, but with marketing and ideological concerns in mind. The analysis of a number of definitions shows which (sometimes unclear) motivations underlie those neologisms. Through examples taken from the fields of comic books and computer science, Section 3 discusses the fact that describing all the relevant semantic features

of terms which do not seem very technical at first sight may require some domain-specific knowledge that a single lexicographer or even a terminologist may lack. In that case, resorting to the diversified expertise of Internet users might prove extremely useful.

1   Tackling the Issue of Neologism Analysis: from professionals to the crowds

In the subsections below, we introduce four crowdsourced dictionaries by highlighting the specific features that are especially relevant for the treatment of the neological phenomena we are interested in. These dictionaries, which are usually less well-known than traditional dictionaries or, at least, less described, do not fit into the usual categories (such as the ones established by Rey-Debove [1971: 19-37] or Pruvost [2006]) and require detailed explanations. Entries from these dictionaries will be compared to those written by professionals taken from the following sources:

– three English dictionaries: The *Oxford English Dictionary* (OED), the *Macmillan Dictionary* (MMD) and the *Longman Dictionary of English Language and Culture* (LDELC). The OED is the dictionary with the most extensive coverage, is very regularly updated and provides fine-grained analyses of polysemy. The MMD and the LDELC are learners' dictionaries and provide their users with explicit information. The LDELC's specificity lies in the cultural background it provides users with, which makes it particularly useful when it comes to a precise understanding of the referents and the treatment of connotation in particular.

– one French dictionary: the *Petit Robert* (PR), a general-purpose, one-volume dictionary, whose online edition is used in this paper to compare the detection of neologisms with what is done by *Wiktionnaire* (WIKTFR) in section 1.2, and to contrast the definitions given in English and in French for referents which are supposedly identical in section 3.

– two term banks: the *Grand Dictionnaire Terminologique* (GDT) and *Termium Plus* (TERM), which are two multilingual term banks developed by Canadian institutions (the Office Québécois de la Langue Française for the GDT and the Translation Bureau for *Termium Plus*), and to which amateur dictionaries will be compared in section 3 for the treatment of some specific terms.

1.1 Main Features of a Selection of Crowdsourced Dictionaries

The adjective used to qualify the dictionaries under study needs to be commented upon. The use of the morpheme *crowd* in *crowdsourced* is particularly debatable. While the term *crowd* seems to relate to a massive amount of contributors, the fact is that only a few people contribute on a regular basis [Sajous and Hathout, 2015]. Moreover, the term *crowd* cannot be applied to some *amateur dictionaries* written by a single author, such as JargonF. Contrasting collaborative dictionaries with professional dictionaries, on the other hand, would suggest that professional lexicographers do not collaborate. Finally, using the term amateur dictionary only makes sense from the lexicographical competences' perspective: an author can be an amateur lexicographer, but an expert of the field he is writing about. Let us now briefly present the main features of the four dictionaries under study: *Wiktionary*, *Urban Dictionary*, JargonF and *Macmillan Dictionary*.

Wiktionary (WIKT) is a dictionary project which relies, just like Wikipedia, on the "wiki principle": any internet user can edit any entry and each change is published at once. The dictionary claims not to be prescriptive, and intends to collect rare and obsolete words, as well as neologisms, taken from the general language or specific languages, and diatopic (regional and national) variants. The neutral point of view is one of the "imperative and non negotiable" founding principles originating from *Wikipedia*. Besides the expected elements of the microstructure that are relevant for the study of neology (etymology, glosses, etc.), *Wiktionary* may use quotes to give implicit hints on connotation and on the semantic features that definitions cannot signal without departing from the neutrality principle [Sajous and Hathout, 2017]. Another element of Wiktionary's microstructure used to provide additional information is the "usage notes" which can give grammatical information but also correspond to the "word choice" section of some other dictionaries. For instance, the usage note found in definition 1 may not explicitly answer the question "should one say *tax avoidance* or *tax avoidance scheme*?" but gives some hints on the subtle difference between the two words. [1]

The second dictionary under study, *Urban Dictionary* (henceforth UD), "started as the anti-dictionary, a parody of dictionary.com. Today it's not just a parody: Parents and teachers use it to understand the next generation" according to its founder Aaron Peckam quoted by Gao [2012]. The way people

---

1. All online dictionaries were consulted in May 2018, unless otherwise specified.

> **tax avoidance** (*uncountable*)
> 1. The legal exploitation of tax rules to minimize tax payments.
>
> **Usage notes**
> While *tax avoidance,* the minimization of tax by careful reading of tax rules, has traditionally been regarded as legal and legitimate, the issue of tax avoidance *schemes* came into public debate (for example in the UK and Australia) during the latter part of the 20th century. Consequently, some such schemes are considered unacceptable, and may not remove the liability to pay [. . .]

DEFINITION 1: *tax avoidance* [WIKT]

contribute to it is different from the one used by *Wiktionary*. It is a cumulative rather than a collaborative system, since any Internet user can add a new entry, or add a new definition to an existing entry, even if that definition does not correspond to an additional meaning. Internet users cannot edit the definitions written by other authors, but can only vote for or against it. For a given entry, the various definitions are displayed according to a decreasing relevance score, which corresponds to the difference between the number of positive and negative votes. UD and WIKT are also poles apart regarding their editorial policies: in UD, personal points of view are favored, as explained by A. Peckam: "Every single word on here (*sic*) is written by someone with a point of view, with a personal experience of the word in the entry". [2] Often labelled a *slang dictionary*, UD does include "slang" words (in the broadest sense) taken from pop culture and subcultures. It also contains terms from specialized fields such as *pseudo-code* (computer science), *hypernym* (linguistics), as well as general language words. As shown by Damaso and Cotter [2007], UD has also become a sort of entertaining place, where Internet users sometimes create entries for made-up terms, or even an outlet for the frustration of the users who do not hesitate to resort to name-calling, as some complain about. When browsing the dictionary, one is also likely to come across an entry with bawdy references. Finally, some metalinguistic comments are also found there, in which contributors criticize what they consider deviant usage and try to impose or contradict some semantic features, or try to impose or ban the use of some words.

Let us now say a few words about the third dictionary under study, ***JargonF***, whose subtitle, "dictionnaire d'informatique francophone", is more appro-

---

2. StreetSmart: *UrbanDictionary. New-York Times*, July 5 2009.

priate for the website than its main title, *Le Jargon Français*. This dictionary, which has been compiled by a computer professional from the University of Rennes, France, since 1994, contains over 15,000 entries. Even though this dictionary provides French definitions for French words, we found it relevant to include it in our study for four reasons. First, a large proportion of French computer science terms are borrowed from English, and it is interesting to see how many of the original semantic features are kept or modified in the French term. Second, there are a number of features that stand out given the fact that there is a single author: the high number of entries, the subtleties of some descriptions which explicitly point to the cases of homonymy or polysemy, the original and up-to-date meanings, and sometimes the deviant usages. Third, this specialized dictionary is technically accurate. Finally, the author's point of view is always clearly expressed: his own positions are always humorously signaled (for instance he makes it clear that he favors freeware or explicitly criticizes the lack of knowledge of most journalists). Comparing *JargonF*'s (3) and Wiktionary's (2) definitions of *micropayment* allows one to better understand the *raison d'être* of *JargonF*'s mockery:

> (*economics*) A financial transaction for a very small amount of money

DEFINITION 2: *micropayment* [WIKT]

> [banque] Paiement d'un montant relativement faible, pas plus d'un neuro (*sic*) en règle générale. Les systèmes sûrs les autorisant sont rares tellement les banques sont gourmandes : elles voudraient nous faire croire qu'une transaction leur coûte très très très cher...

DEFINITION 3: *micropayment* [JargonF]

When reading (3), one may wonder why a word was coined for the payment of very small amounts of money and not for higher ones. By blaming a banking practice that seems to him unwarranted, *JargonF*'s author implicitly sheds light on an additional piece of information: the fact that there is a tax on this type of payment.

The last dictionary under study, the ***Macmillan Open Dictionary*** (henceforth MMOD), is the crowdsourced dictionary that was launched by Macmillan back in 2009, and which allows Internet users to submit definitions for new entries, which will later be edited by professionals. According to the website, there have been 4,000 additions so far, half of which have been promoted

to regular entries. Even though the *Macmillan Dictionary* (MMD) is not the only professional dictionary that offers a crowdsourced section, the fact that the definitions written by Internet users (and signaled as such) can be accessed through the same search bar as the ones written by trained lexicographers is noteworthy. This is an acknowledgment of the competence of laypersons that goes beyond the one that James Murray – the *Oxford English Dictionary*'s chief editor – granted the amateurs that participated in his reading program – a sort of *avant-garde* snail mail crowdsourced lexicography that was launched as early as 1879.

Lexicographers from the MMD control the relevance of the entries suggested as additions by Internet users and make sure the writing of the definitions is in line with the dictionary's editorial policy ; as a result, they are able, to some extent, to extend its nomenclature, to shorten the inclusion timespan (just like *Wiktionary*), to benefit from the diversified expertise of its various contributors (as in a sort of multi-field JargonF), while avoiding UD's pitfalls.

1.2   Identifying and Recording Neologisms

While the automatic detection of formal neologisms is not much considered an issue any more, it seems semantic neology still cannot be dealt with automatically satisfactorily. Even though the former task is easier than the latter, it remains complex. Beyond the technical and methodological difficulties that cannot be detailed here for lack of space, automatic detection depends on the availability of a large and diversified enough diachronic corpus. Moreover, statistical methods make it possible to detect neologisms only a while after they have first appeared [Falk et al., 2014]. As mentioned earlier, professional dictionaries cannot be used to detect neologisms (whether they use a diachronic label or not, for that matter). In fact, as underlined by Mortureux [2011], it is just the opposite: the lexicalization process is over once a word has been recorded in the dictionary. She also writes that the lexicographic discourse about the new lexeme provides explanations about the way it is integrated into the lexical system by mentioning other lexemes the neologism is now structurally related to. For a neologism to be included in *Wiktionary*, neither the ever-increasing use nor the complete analysis of the word is a prerequisite. A dictionary entry is being published while being built: as soon as a neologism is identified, a webpage including the headword and the part of speech is created. The definition, pronunciation, etymology and lexical semantic relations are added later. Such a continuous publication mode fosters the early inclusion of neologisms in such a crowdsourced dictionary. Hathout

et al. [2014] have shown that Wiktionary covers general-purpose French language quite extensively; not only does its unrivaled nomenclature include rare, technical or obsolete words, but it also covers vocabulary from diversified corpora, which is clearly an advantage over other resources. We wondered if, for English, the discrepancy between crowdsourced and professional dictionaries was the same as for French regarding their nomenclature and the time it takes for a neologism to be included.

Our question stems from the difference in evolution between English and French lexicography. While the former started to rely on corpus linguistics quite early and benefited from the collaboration between lexicographers and linguists (some specializing in computational linguistics), both from the private and the academic sectors, the same cannot be said about the latter, as obviously stated in the titles of two of Pierre Corbin's papers: "La lexicographie française est-elle en panne ?" [Corbin, 1998] and "Quel avenir pour la lexicographie française ?" [Corbin, 2008]. In France, it seems the border between private publishing houses and academic research cannot be crossed, and the somewhat ideological rejection of corpora by some influential linguists [3] has had long-lasting effects.

For the year 2017, we compared the new entries found in the OED and *Wiktionary* on the one hand, and the ones found in the online *Petit Robert* and in *Wiktionnaire* on the other hand. This comparison can seem disproportionate at first sight. The OED is a multivolume dictionary ; with a 600,000-headword nomenclature, it aims at recording the whole lexicon of the English language. The PR is a single-volume dictionary whose nomenclature includes "300,000 words and meanings" (electronic version) or "60,000 words and 300,000 meanings" (paper version). The reader may wonder why we did not use the electronic version of the *Grand Robert* instead, since it includes "100,000 words and 350,000 meanings". Unlike the OED whose complete list of new entries is published every time the dictionary is updated, [4] the Robert publishers only mention a few in occasional press releases. As a consequence, our study is based on the neologisms in the PR that have been manually and carefully identified by Martinez [2009], who does so on a yearly basis – which finally turns out to be a better choice for the analysis of neology since it is updated more often than the GR.

---

3. When speaking about the philological approach used for the *Trésor de la Langue Française* and then for the *Collins Cobuild*, Rey [1995] talks about the "scientist frenzy of machine-readable corpora that can be found in Great-Britain especially" (our translation).

4. http://public.oed.com/the-oed-today/recent-updates-to-the-oed/

As for *Wiktionnaire* and *Wiktionary*, their nomenclatures have been extracted from GLAWI [Sajous and Hathout, 2015; Hathout and Sajous, 2016] and its English equivalent ENGLAWI, two machine-readable dictionaries designed by converting the crowdsourced dictionaries into XML format. As part of a homogeneous treatment, the distinction between simple and compound lexical items only relies on the presence or absence of a blank character. The forms that were taken into account are the lemmas of the lexical words (common nouns, verbs, adjectives and adverbs); for the OED, we only studied the main entries (without the subentries); finally, we limited ourselves to the PR's main entries and compounds (and excluded its "hidden" entries). For the year 2017, the number of additions found in *Wiktionary* is especially high (over 84,000). As a comparison, the average number of additions for *Wiktionary* and *Wiktionnaire* is the same for the 2006-2016 period: +25,000 without the proper nouns, and +35,000 with them. The differences in the OED / *Wiktionary* and PR / *Wiktionnaire*'s coverage shown in Table 1 could be accounted for by the difference in size between the OED and the PR's nomenclature, or by the fact that the number of new entries in *Wiktionary* is nine times as high as in *Wiktionnaire* for 2017.

Wiktionnaire entries		PR		PR entries		*Wiktionnaire*	
		Present	Missing			Present	Missing
mono	6,915	563 (8 %)	6,352 (92%)	mono	112	104 (93 %)	8 (7%)
poly	2,919	228 (8 %)	2,691 (92 %)	poly	36	12 (33 %)	24 (67 %)
TOTAL	9,834	791 (8 %)	9,043 (92 %)	TOTAL	163	128 (79 %)	35 (21 %)

Wiktionary entries		OED		OED entries		*Wiktionary*	
		Present	Missing			Present	Missing
mono	72,703	29,792 (41 %)	42,911 (59%)	mono	264	198 (75 %)	66 (25%)
poly	11,565	1,954 (17 %)	9,611 (83 %)	poly	55	34 (62 %)	21 (38 %)
TOTAL	84,268	31,746 (38 %)	52,522 (62 %)	TOTAL	319	232 (73 %)	87 (27 %)

TABLE 1: Nomenclature Comparison: New Entries for 2017

However, how neologisms are detected and selected for inclusion in the French private sector dictionaries is somewhat dubious. Even though Sommant [2000] describes this process as methodical for Larousse (for the 1988-2000 period), she also mentions a sort of "neologism hunting process" relying on the lexicographers' "flair" and "sixth sense". Since there is no published information other than what is found in the prefaces to the dictionaries and the yearly press releases, only a metalexicographic approach allows us to guesstimate how French lexicographers proceed, and there is no evidence that the methods currently used are more innovative. Martinez [2009] has shown that there is

no coherent pattern regarding those dictionaries' updates. What the PR's press release focuses on is the addition of recent words, such as *hackathon, youtubeur, fablab* in the "computing and multimedia" category, *fixie* and *aquabike* in the sports category, but a closer look at the additions raises a number of questions: why was such word not in the dictionary, and why was it added in 2017? This is the case, for instance, for *chamallow, feignasse, dégun, mort-vivant, pogoter*, etc. For the words added in 2017 in the OED (or PR, resp.) and already recorded in *Wiktionary* (or *Wiktionnaire*, resp.), we checked for how long they had been recorded in the crowdsourced dictionaries – this is what is summed up in the boxplot found in Figure 1. We can notice that the median period of time elapsed between inclusion in *Wiktionary* and in the OED is 4 years long, but that it is almost twice as long (7.5 years) when we compare *Wiktionnaire* and the PR. Another striking fact is that 25% of the words added in the OED in 2017 were recorded in *Wiktionary* the very same year. The next 25% were recorded with a 1- to 4-year delay. The upper limit of the first quartile of the PR (0-5 year period) is beyond the OED's median deviation.
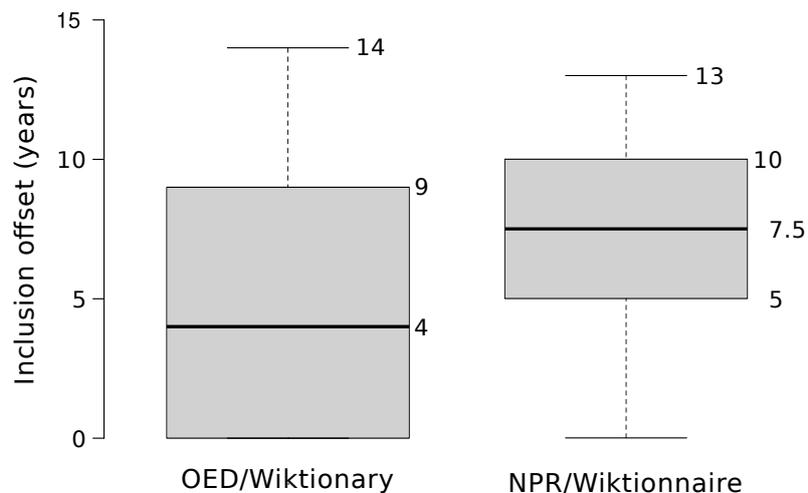

FIGURE 1: Inclusion date: Discrepancy between Professional and Crowdsourced Dictionaries

The maximum deviations (13 and 14) correspond to the time elapsed since the birth of the collaborative dictionary. It would be interesting to check if those deviations are on the rise in the next few years ; one can assume that a word added in *Wiktionary* nowadays is more likely to correspond to a real neologism than a decade or so ago, when the dictionary was catching up on the "core lexicon" while trying to record the then neologisms at the same time. Another interesting study could consist in a qualitative analysis of a sample

of those additions in order to identify specific trends in the type of neologism recorded in a given dictionary.

As a first conclusion, this study shows that *Wiktionary*, thanks to the size of its nomenclature and the will of its contributors to rapidly record neologisms in it, can be a useful complementary tool for the OED's lexicographers. The so-called exclusion corpus – as dictionaries are usually considered – could thus prove to be a very efficient tool for neology watch.

1.3   Motivation Analysis: Spontaneous vs. Planned Neology

The linguistic processes underlying "spontaneous" neology have been descri-bed, among others, by Campbell [1998: 254-279]. We will focus on linguistic planning, i.e. the conscious effort of some speakers to urge a speech commu-nity to use new lexical units or to modify existing lexical units. Díaz Hor-mingo [2012] focuses on euphemisms and underlines the fact that the typo-logical features that are commonly established to describe them (e.g. the dis-tinction between denominative or referential neology vs. stylistic or expres-sive neology) are inadequate. For the analysis to be thorough, Díaz Hormingo considers that the extralinguistic motivations underlying the euphemistic crea-tions need to be examined. To follow suit, we strongly believe that the analysis of motivation should be at the core of the analysis of any neologism analysis, from the moment that the neologism under study is due to language planning. This is something seldom found in professional dictionaries, either for lack of space, or because lexicographers would be faced with expressing their point of view. Turning then to amateur dictionaries can prove useful to answer the following questions: Who initiated the change? What was the change for? Is its motivation hidden or openly displayed, transparent or opaque?

When official institutions (such as OQLF in Quebec or DGLFLF in France) are in charge of such language enrichment, by replacing borrowings by lexi-cal creations which are to be published (for instance in the French *Journal Officiel*), the answer to the three questions is rather straightforward. However, in some cases, language planning goes unnoticed – be it official or not. This is the type of initiative that Ager [2001] describes by analyzing both the motiva-tion and the historical context. Language planning can relate to the language being used ("discourse") but also to dictionary-making – Ager reminds rea-ders that some dictionaries were compiled on royal request. In a similar vein,

Farina [2016] quotes the example of a dictionary whose etymological information was deliberately falsified in the pre-Soviet era in order to exacerbate jingoism. In a text entitled "Politics and the English Language" published in 1946, George Orwell warned against the risks of language being manipulated for political purposes: "if thought corrupts language, language can also corrupt thought", with a sense of foreboding about how the language was to be corrupted in the totalitarian regimes of the time. According to Krieg-Planque [2012], the book *LTI, Lingua Tertii Imperii* written by the philologist Victor Klemperer and published in 1947, during the Second World War, was to confirm Orwell's intuition by describing how the Nazi regime brought a change to the German language and reduced its combinatory possibilities in order to reduce the capacity for collective and individual thought.

While these are extreme cases, the linguistic changes initiated by the political and economic powers that Farina [2016] calls "Top-Down Linguistic Innovation" are still relevant. The very recent example she gives is the following: when Crimea was annexed by Russia, Vladimir Putin used the word natsional-predatel ("national traitor") to refer to the Russians that were opposed to the annexation. Historically speaking, this word has been used to talk about those who collaborate with the occupying enemy – the only thing being that there is no foreign enemy currently occupying Russia. It seems rather obvious that Putin is aware of that when he addresses Parliament, but that he relies on the official media to spread the use of the word among the population. According to a paper published by the *Washington Post* on December 15, 2017, the Trump administration has banned the use of the words *vulnerable, entitlement, diversity, transgender, fetus, evidence-based* and *science-based* at the Department of Health and Human Services. Such censorship is likely to lead to the use of some neological substitutes. On April 15, 2018, in an interview given to the newspaper *Mediapart* and the TV channel *BFMTV*, French president Emmanuel Macron tried to convince a journalist to talk about *optimisation fiscale* while the questions he was asked actually dealt with *évasion fiscale*.

Even though the issue at stake in the three examples we have just mentioned is not exactly the same, such practices – be they a case of discrepancy between the recorded meaning and the way it is actually used, or cases of censorship or euphemism – are likely to spread and to end up in dictionaries when they are updated. Seldom do dictionaries provide an answer to the question: "who coined this phrase / who used a new meaning for that word and why?", either because it is not part of their editorial policies or because

there is no straightforward answer. Once the use of a lexical unit is settled in discourse, its history can be hard to trace. Another difficulty is underlined by Ager [2001], who identifies three distinct planning instances: individuals, communities and states. This classification needs to be refined, since some multinational companies sometimes have a state-like behavior, and there now are some supranational organizations, such as the Commonwealth or the EU. Even in the case where the editorial policy allows it, it might be hard to identify who lies behind the neologism and what his/her intention was in coining it. There is one noteworthy exception, though, with the *economic migrant* entry of the MMD (4) which clearly states who uses this phrase (governments) and why (namely, to distinguish a *migrant* from a *refugee*).

> someone who goes to a new country because living conditions or opportunities for jobs are not good in their own country. This word is used by governments to show that a person is not considered a refugee (= someone who has been forced to leave their country for political reasons).

DEFINITION 4: *economic migrant* [MMD]

Such a "departure from lexicographic neutrality", which Atkins and Rundell [2008: 427-430] call *editorializing*, is seldom found in general-purpose dictionaries, and more often in culture-oriented dictionaries such as the LDELC. Using such dictionaries which are originally aimed at language learners could be highly beneficial to native speakers in order to better grasp the meaning and the connotation of a given word and to understand the reasons underlying any linguistic change. In the case of the words *migrant / economic migrant / refugee*, reading about controversial debates led by journalists [5] or about the analysis of metadiscursive statements related to that topic (see Calabrese [2018]), who shows that the clarification given in (4) is far from being superfluous. Refusing to use some words that are available or using "the right word" is only seemingly insignificant, since some societal and ideological issues are at stake. The two positions (boycott or use of the recorded meaning) are often found in glossaries written by various groups or individuals with a pedagogical or an activist aim, as explained by Krieg-Planque [2012]. Thus, the "petit guide *Lutter contre les préjugés sur les migrants*" written by the

---

5. See for instance the article "The readers' editor on... the semantics of migration" published in *The Guardian* (08/16/2015), *Al Jazeera*'s article "Why Al Jazeera will not say Mediterranean 'migrants'" (08/20/2015) or the one published by the Italian radiostation *Life-Gate* entitled "Don't call them migrants, call them refugees" (09/29/2015).

CIMADE [6] chose to define the French equivalents of the terms *migrant, economic migrant, refugee* and *exilee* in order to help the users use them with full knowledge of the facts.

Lecolle [2012]'s study focuses on the "sentiment linguistique profane" (which corresponds more of less to folk linguistics), and more particularly on the ability of speakers to detect linguistic change. Searching markers inspired by the ones used by Lecolle (i.e. *euphemism, not to be confused with, politically correct, newspeak, weasel word*, etc.) in UD reveals that the contributors to this dictionary are sensitive to neology and analyze the motivation underlying the changes they have detected. Internet users inform readers about the discrepancies between the recorded meanings and the distorted usages in discourse, and often criticize them. A prime example is the incorrect use of the word *semantics* when what is at stake is form and not meaning – something named "sémantique blabla" and described by Le Draoulec et al. [2014]. One of the contributors describes the word as follows:

> The study of the meaning of words.
> Often misused when quibbling about something someone said. In that context, the statement "That's only semantics" would be more aptly phrased as "You're just splitting hairs on word meanings." [...] The very concept of semantics is frequently disparaged by wishy-washy passive-aggressives who refuse to be accountable for their careless use of language or their deplorable lack of education.

DEFINITION 5: *semantics* [UD, def. #3/7, 2009]

The contributor who wrote this definition provides the "real" meaning of the headword ("the study of the meaning of words") together with the wrong usage that he/she has noticed ("misused") which he/she blames on some speakers' lack of knowledge. Such adjusting notes can be used in UD to criticize "lexical manipulations" (see section 2) but also to correct or disprove the semantic features that have been assigned to some lexical units (see section 3).

1.4 Specialized Neologisms: a Thorny Issue

When a formal neologism meets a denotational need due to the existence of a new concept (e.g. *e-cigarette*), there is no motivation issue at first sight. What

---

6. A French association supporting foreigners, especially regarding their legal rights.

remains to be done then is to identify the exact meaning of the neologism. Terminological neologisms, or *neonyms*, are said to be characteristically monoreferential and unequivocal, semantically stable, with a neutral connotation and no synonym. However, since there is a lot of going back and forth between specialized discourse and the general language, the meaning of terms tends to be altered within various speech communities. One reason for this is that defining terms does not only rest on terminologists: it is often a lexicographer's task when compiling general-purpose dictionaries. In that case, the issue often raised is that the main difficulty the lexicographer is faced with when describing the term is how to adapt the definitions to the audience, but little is said about how challenging it can be for him/her to grasp the exact meaning of a term. The distinction stated by Bowker [2003] between several levels of expertise among the target audience (i.e. true experts, semi-experts and non-experts) may also apply to lexicographers. When a "general" lexicographer's skills or documentation cannot provide him/her with a satisfactory answer when it comes to deciphering a particularly abstruse neologism, he/she can always turn to experts. We may wonder if this is actually what he/she does, and whether it bears fruit or not. Čermák [2003] suggests finding, beyond the usual documentation, "further pragmatic information about use, clarification, or definition of an item". Béjoint [1998] adds:

> No amount of context can specify all the semantic features necessary for an adequate definition [...] If one wants to be sure to capture all the semantic traits of scientific or technical words, the only option is to ask specialists of the domain to define them.

The level of specialization of terms in some fields, such as *diffusion porosity, compression crack* or *attrition mill* in the field of powder metallurgy is such that a lexicographer cannot rely on his/her experience or intuition. Some terms which seem less technical may turn out to be just as hard to define satisfactorily. Since some fields are less technical than others (e.g. comics), some others are intertwined with everyday life (e.g. computing) or since within a given field, some terms are less specialized than others (e.g. *hacker* vs. *code injection* in computer security), the lexicographer may do without the help of experts or the analysis of a large enough number of contexts taken from a corpus. The boundary between specialized languages and the general language is often blurred and their intersection provides fertile ground for neology, as shown by Meyer and Mackintosh [2000]. For instance, when a term is "determinologized", it can have an additional meaning in the general language, which can also consequently alter the initial specialized meaning. The concepts themselves may also evolve over time, which will induce semantic shifts. Section 3 shows that, in that particular case, to fully understand

a word or concept that seems clear at first sight, it might be necessary to first see where it stands within the field's specific culture before turning to what Pruvost [2005], following R. Galisson, calls "lexiculture" (i.e. shared common lexical knowledge). A concordancer or an expert might be insufficient ways to penetrate that specific culture. Conversely, a field or subculture expert may have enough background to define a concept from his/her own field. One of the main features of crowdsourced dictionaries is precisely the diversity of its contributors, as far as training, social and geographical origins, tastes and interests, occupations, etc. are concerned. Each amateur lexicographer is a potential specialist for a given field. In Wiktionary, this shows through the completeness of the coverage of many technical fields. It is rather different in UD, which is characterized by its militancy and identity-building values. It is occasionally used for "semantic adjustments", where contributors can disprove some semantic features or connotations wrongly assigned to a headword. Such definitions can be spotted thanks to the pattern: [entry] + negative form of "be", e.g. *"feminists/lesbians **are not** man-haters"*, *"vegetarians **are not** elitist"*, etc. Some beliefs that are much less ideologically oriented are also discredited: *"Integration **is NOT** the reversal of differentiation"*, *"longboards **are not** necessarily longer than freestyle skate-boards"*, *"Lager **is NOT** an ale"*. These corrections correspond to what contributors perceive as generally agreed societal connotations, or can also be a criticism of the definitions found in some traditional dictionaries – since the latter, as stated by Dubois and Dubois [1971: 99-104], mirror the cultural norm that is in keeping with the ruling classes' ideology. UD contributors, who potentially belong to various communities and share values of "fringe cultures", seem to respond to that "mainstream" ideology through their definitions. However, the spectrum of UD contributors is broad and this offbeat dictionary can itself be considered as a mainstream product. Farina [2005] shows how dictionaries still are – intentionally or not – a vehicle for stereotypes, and thus contribute to spreading and reinforcing them, in particular the sexist ones. UD is no exception to that, and even includes self-criticisms, as can be seen in the definition 6 for *sexist*.

> [...] Attitudes, conditions, or behaviors that promote stereotyping of social roles based on gender.
> *Urban dictionary is a great example of sexist attitudes and phrases used in culture today.*

DEFINITION 6: *sexist* [UD, def. #2/7, 2008]

Adjustments can relate in UD to words that are not neologisms (e.g. the previous *feminist* and *vegan* examples) or to more recent words. The OED's update released in March 2018 lists many additions related to gender and sexual identity [7] issues, including *transgendered*. There are 5 definitions for that word in UD, the oldest one going back to 2006. It was actually used to define the word *genderqueer* (def. 7) by contrast – a word that was first recorded in 2004 in UD and in 2011 in the OED.

> Any position in a wide variety of gender identities, spanning the spectrum between male and female. A person who is Genderqueer **is not** transgendered, though the option is open to transition. Being genderqueer has no bearing on sexual identity or orientation.

DEFINITION 7: *genderqueer* (bold emphasis added)
[UD, def. #2/16, 2006]

Section 3 will show that such negations are either a response to other existing definitions or a dismissal of the relevance of a semantic feature for a given word ("X is not defined it terms of").

2 Lexical Manipulations

2.1 Euphemisms: from Political Correctness to Propaganda

Díaz Hormingo [2012] draws a distinction between euphemistic substitutes "that are motivated for both the speaker and the hearer" and those "that are motivated for the speaker but not the hearer". The former, which could be called "propriety euphemisms", address the need to avoid crude lexical items (linguistic taboos such as *the f-word* for *fuck*) or relate to some realities (cultural taboos such as *powder room* for *toilet*). The latter aim to disguise reality and to divert the co-speaker's attention. This is particularly true of examples taken from political speeches, such as the ones picked by Crespo-Fernández [2014] in the regional British press. Speaking about *issues of misconduct* allows a local representative to play down the violence that was perpetrated against night-time revelers in order to preserve his city's night-time economy. The linguistic processes that are used are varied – understatement, circumlocution, use of the passive voice that suggests no one assumes liability, just

---

7. `http://public.oed.com/the-oed-today/recent-updates-to-the-oed/march-2018-update/reelase-notes-formal-language-sexuality-gender-identity/`

like the underlying motivations – from face-saving to deception. Unlike the "propriety euphemisms" mentioned above, which are very likely to be found in a dictionary, the ones on which Crespo-Fernández's study focuses are only found in discourse and, as a consequence, are seldom described in dictionaries.

In *Wiktionary*, 934 definitions are labelled euphemistic. Among those, 33 are phrased as follows: "the word X", where X is a synonym of the headword being defined. These are mostly entries which follow the *s-word* and *f-word* pattern (27 definitions in total for 18 initial letters). In some cases, the comment "regarded as a vulgar or taboo word" is included in the definition, which explicitly accounts for the need for a substitute. In some other cases, the definitions are phrased as follows: "synonym of X", most of the time to define a substitute word for *WC* or *toilet*; *powder room* is thus defined as a euphemistic synonym for *ladies' room*, which is itself labelled *euphemistic* and defined as "a public *lavatory* intended for use by women". Even though the word being substituted, *lavatory*, is mentioned in the definition, the reason for which it should be avoided – an implicit cultural taboo that is deemed universal – is not mentioned. However, the so-called universality of that taboo can be discussed, and the use of words such as *WC* or *toilet* (and all the more so of *ladies' room*) might not be such an issue in some cultures, as noticed by Frank McCourt in his autobiographical novel *'Tis* where he writes about being faced with American English as a young Irish immigrant:

> If you want a WC or a lavatory you have to say bathroom even if there isn't a sign of a bath there. And no one dies in America, they pass away or they're deceased and when they die the body, which is called the remains, is taken to a funeral home.

The definitions of the above-mentioned euphemisms in a cultural dictionary for learners (LDELC) are also revealing:

n euph a women's public TOILET in a theater, hotel, restaurant, big shop etc.

DEFINITION 8: *powder room* [LDELC]

AmE–n a women's TOILET – compare GENTS; see TOILET (USAGE)

DEFINITION 9: *ladies room* [LDELC]

While definition 8 is the only one with a euphemism label, definition 9 has a cross-reference to the usage note for toilet which we only partly copy in definition 10. Despite quite a detailed explanation, some elements are still

> In British English **toilet** is generally acceptable, but **lavatory** and **WC** (becoming old-fashioned [...]) are also used. **Public conveniences** is the formal expression [...] these are also called **the gents** or **the ladies**. In American English **bathroom**, **restroom** and **washroom** are commonly used for **toilet**, and **john** is a common informal word.

<div align="center">DEFINITION 10: *toilet* [LDELC]</div>

implicit: toilet is said to be "*generally* acceptable", but in which cases is it *not* acceptable? Should *lavatory* and *WC* be substituted only because they are old-fashioned? Are the substitutes "commonly used for toilet" in American English used because of the words being substituted are old-fashioned or taboo? Krieg-Planque [2007] studies what she calls "les jugements d'euphémisation" ("euphemizing judgment"), which is described as a metadiscursive operation that consists in explicitly referring to a certain type of phrasing as euphemistic – a sort of assessment of the discrepancy between the word and the thing it refers to. Dictionaries, she says, rubber stamp the (objective and universal) existence of that discrepancy. She strongly believes it is the use of a word *in a given situation* that makes it a euphemism *for someone at some point*.

Adjectives coined by adding the *-challenged* or *-impaired* suffixes to adverbs to refer to some handicaps (e.g. *physically challenged*, *visually impaired*) are a case of euphemisms where there does not seem to be any discrepancy between the word and the thing it refers to ; the only matter at stake is connotation and, in the pattern "the word W is a euphemism for word Y", saying "for whom" is then optional. The definition for *physically challenged* in *Wiktionary* is indeed:

> (*euphemistic*) Having some physical disability; disabled.

<div align="center">DEFINITION 11: *physically challenged* [*Wiktionnaire*]</div>

and the one found in the OED for the headword *'challenged* adj. (under *'challenged* v.) reads:

> euphem. With prefixed adverb, also occasionally forming nouns.
> a. orig. and chiefly N. Amer. Of a person: disabled or handicapped, esp. physically.

DEFINITION 12: *'challenged* [OED]

Even though the euphemism status can be perceived as universal, one may wonder why adjectives such as *disabled* or *blind* can be considered an issue. There is no clear answer to that question in the dictionary ; in just the same way, the extract from the *NY Times* used as an example (13) in the OED does not provide any explanation for Kennedy's preference.

> 1985 N.Y. Times 20 Apr. I. 26/3 The disabled skiers, whom Mr. Kennedy prefers to call 'physically challenged', achieve speeds on difficult runs that would be daunting to most competitors. [OED]

DEFINITION 13: *disabled* (example) [OED]

Excessive political correctness is sometimes mocked by ironic coinages such as *parentally challenged* which is labelled euphemistic in *Wiktionary* and defined as "Lacking one or both parents, or having parents who are inadequately supportive or caring". Understanding what is proper is not trivial for a learner of another language and culture: faced, just like Frank McCourt, with euphemisms such as *pass away* or *bathroom* labelled as euphemisms for *die* or *WC* in the dictionary, should learners assume that such words are improper? Nothing in the dictionary allows him/her to decide whether it is the word being substituted which is a taboo word (not labelled as such, but then why would that be a euphemism?) or if it is the referent that relates to a cultural taboo. The expression *between jobs*, which is more recent than the "universal" ones just mentioned, is defined as follows in three crowdsourced dictionaries :

> (*euphemistic*) Unemployed

DEFINITION 14: *between jobs* [WIKT]

> temporarily unemployed

DEFINITION 15: *between jobs* [MMOD, submitted on 09/09/2015]

> A clever and discreet way of saying your unemployed. you've had a job, not currently working, but will be working in the near future

DEFINITION 16: *between jobs* [UD, def. #1/1, 2007]

Once more, there is no explicit explanation as to why the word unemployed should be avoided, even though the definition found in UD seems to imply that one should conceal being unemployed. By reading the answer below to the question "What is the meaning of the phrase 'in between jobs'?" asked on a forum (quora.com), one can assume that, for unemployed people to look respectable, job-seeking is a prerequisite:

> [answered 19 Mar 2018 by a former Business Analyst (Retired)] Strictly speaking, the phrase means that the person who says it has recently left a job and is planning to commence another job. It is often used as a euphemism for 'unemployed', although it would be the truth provided the person who says it is seeking employment.

In French, for that matter, *personne en recherche d'emploi* ("jobseeker") is a euphemism for *chômeur* ("unemployed person"). Guilt-inducing social pressure implicitly accounts for the widespread use of that expression, but drawing such an inference does not necessarily immediately spring to mind.

"Lexical cosmetics" is the term used by Nida [1995] to refer to the fact of using words or phrases in a way where the positive features of the concepts represented by the concepts are highlighted, while their negative features are minimized. The example he gives is *pro-life* and *pro-choice*, which are substitutes that are more positive and acceptable than *anti-abortion* and *pro-abortion*. It seems to us than there is more to that substitution than a shift from the negative connotation of *abortion* to the positive connotation of the prefix *pro-*. Fighting for some values implicitly means depriving one's opponents of those values, or even associating them with the opposite values. Accordingly, a *pro-life* opponent is believed to be implicitly *anti-life*. This is what is in the minds of the Internet users who wrote the following definitions 17 and 18.

> A term invented by anti-abortionists to refer to being anti-abortion. Everyone is pro-life you idiot, what you are is anti-abortion.

DEFINITION 17: *pro-life* [UD, def. #9/34, 2004]

> Politically correct label that applies to people who oppose a woman's right to chose (*sic*). The choice of words, 'pro-life', suggests that the people who don't agree with them are 'pro-death' and therefore bloodthirsty babykillers.

DEFINITION 18: *pro-life* [UD, def. #12/34, 2004]

The definitions found in LDELC for *pro-life* (19) and *pro-choice* (20) can be disturbingly implicit: we do find the label *euph.*, but, in the same way as Krieg-Planque [2007] asks "for whom is the word euphemistic?", the reader may ask "which word is the entry a euphemism for?".

> *euph* opposed to ABORTION

DEFINITION 19: *pro-life* [LDELC]

> *euph* favouring ABORTION being available to those who want it. In the US, people who are pro-choice often LOBBY Congress and walk in DEMONSTRATION – compare PRO-LIFE; see also Cultural Note at ABORTION

DEFINITION 20: *pro-choice* [LDELC]

More explanations are provided in the cultural note for *abortion* (21), but the extralinguistic reasons that led to the suppletion of *anti-abortion* by *pro-life* are not made explicit.

> Abortion has been legal in the US since 1973 and in the UK since 1967, although people in both countries have very strong opinion about it. People disagree about whether it should stay legal, and about whether it is morally right. Some people are PRO-CHOICE and believe that a woman has the right to choose whether or not to have an abortion. Other people are PRO-LIFE or anti-abortion, and believe that an unborn baby has the right to be born. They believe that abortion is murder. – see also ROE VS. WADE

DEFINITION 21: *abortion* (cultural note) [LDELC]

The segment "other people are PRO-LIFE or anti-abortion" seems to show that the two expressions mean the same thing, but does not include any contrastive element. Since the dictionary does not have an entry for the second term,

the reader has no idea whether it is derogatory or not. Providing information in a very implicit manner (as we have seen in definitions 8 to 21) is very common in dictionaries aimed at adult native speakers; not much so in learners' dictionaries. According to Girardin [1979], the excess of implicit explanations can be considered as censorship, for instance when the reader needs to infer too many elements.

When pushed to its limits, Nida's lexical cosmetics becomes a matter of lexical usurpation with a manipulating purpose. This is what is described by Krieg-Planque [2015], who studies frozen structures as a strategy to build authority in discourse. Traces of vocabulary originating from political, economic or managerial newspeak can be found in dictionaries without any critical distance. She provides examples from the socio-political vocabulary which, by being excessively repeated, end up being established and taken for granted, thus making any deviating or opposite ideology invisible.

Back in 1946, in *Politics and the English Language*, Orwell already denounced euphemizing phraseology:

> Defenceless villages are bombarded from the air, the inhabitants driven out into the countryside, the cattle machine-gunned, the huts set on fire with incendiary bullets: this is called *pacification*. Millions of peasants are robbed of their farms and sent trudging along the roads with no more than they can carry: this is called *transfer of population* or *rectification of frontiers*. People are imprisoned for years without trial, or shot in the back of the neck or sent to die of scurvy in Arctic lumber camps: this is called *elimination of unreliable elements*.

The parallel that can be drawn with what was written 66 years later by Díaz Hormingo [2012] is striking (and stresses the relevance of Orwell's thoughts):

> [...] the present economic crisis, the first Gulf war (1991), the sinking of the tanker Prestige (2002), the Iraq war (2003) or certain terrorist attacks have given rise to appearances in the media by public figures having social responsibility coining lexical units such as economic deceleration for crisis, negative growth for decreased productivity, recession for long-lasting serious economic situation, redundancy for mass sackings, price adjustment or revision for price rises, creditors meeting for temporary receivership, allied attack for war, collateral damage for civilian casualties, humanitarian aid for logistical support or military support [...]

As far as the military field is concerned, one may compare the definition for

collateral damage (missing in the OED) in the MMD (22), *Wiktionary* (23) and UD (24).

> ordinary citizens who are killed during a war. This word is used especially by military officers.

DEFINITION 22: *collateral damage* [MMD]

> 1. (*military, euphemistic*) Damage to civilian property or civilian casualties that are the unintended result of military operations.
> 2. Harm to innocent people that results from policy decisions.
> 3. Unintended victims of an attack targeted at someone or something else.

DEFINITION 23: *collateral damage* [WIKT]

> Military slang for the mass murder of civilians through the use of weapons which are known in advance to be imprecise and/or to cause damage across a large area (e.g. cluster bombs).

DEFINITION 24: *collateral damage* [UD, def #2/10, 2004]

The MMD definition is clear, but the word is not presented as a euphemism. In addition to mentioning "mass murder", the definition found in UD highlights a sense of premeditation (ignoring consequences that can be anticipated). To what extent the three different meanings provided in *Wiktionary* are distinct or overlapping is quite hard to identify. A *euphemistic* label can nevertheless be found in the first definition, and, rarely enough in that dictionary, a point of view is clearly expressed in the second one.

On a different note, human resources are also a field in which authoritative discourse can be built. Let us now compare the various definitions of *restructure / restructuring*:

> to organize something such as a company in a different way so that it will operate better. Derived word: restructuring

DEFINITION 25: *restructure* [MMD]

> Literal meaning: to rearrange the structure of something. Real meaning: to sack lots of workers. A euphemism used by bosses to cover up the fact that they're planning to sack workers or to attack workers in other ways (e.g. worsen pay and conditions).

DEFINITION 26: *restructuring* [UD, def #1/1, 2004]

> Corporate speak for fire or make redundant.

DEFINITION 27: *restructure* [UD, def #1/2, 2005]

UD, taken as a whole, cannot be labelled a militant dictionary. However, definitions such as 24, 26 and 27 show that UD can be considered to some extent as what Krieg-Planque [2015] calls a "tool to deconstruct authoritative discourse".

To conclude this section, one may wonder, when telling the difference between two competing lexical units is challenging, whether that difference is linguistic or conceptual. In *Wiktionary*, the difference between *tax avoidance* and *tax evasion* when defining the terms is straightforward: the former is legal, unlike the latter. In real life, there is no clear-cut distinction between the two ; even though *tax avoidance* is said to be legal, both practices are fought by the OECD (Working Party 11 is entitled *Aggressive Tax Planning* [8]). A usage note in *Wiktionary* explains the subtle difference between *tax avoidance* and *tax avoidance* **schemes**. The portmanteau word *avoison* stems from the difficulty to make a distinction between legal and illegal tax schemes. The words used in French are *fraude fiscale*, *évasion fiscale* and *optimisation fiscale* (or even *évasion fiscale agressive*). The website of the French Economy and Finance ministry includes a note, according to which the terms *évasion*, *optimisation* and *fraude* are closely linked. [9] It even adds that "unlike *fraude*, *optimisation* is legal even if its legitimacy or efficiency may be debatable" ; and that "the definition of *évasion fiscale* is more complex, since it relates both to *optimisation* and to *fraude*".

The very existence of linguistic variants and the vague feeling of euphemization felt by the speaker can be due to conceptual fuzziness. In that particular

---

8. OECD's work on Aggressive Tax Planning: http://oecd.org/tax/exchange-of-tax-information/atp.htm
9. Évasion, fraude, optimisation fiscale : quelles différences ? http://www.economie.gouv.fr/facileco/evasion-fraude-optimisation-fiscale

case, it will be difficult for a general-purpose dictionary, whatever its editorial policy, to record and describe accurately such an opaque and unstable reality, except maybe in a note such as the one given in definition 1: if the semantic fuzziness is described as such in the dictionary, the reader will feel reassured and will not blame himself/herself for misunderstanding.

2.2 Oxymorons

Oxymorons are a stylistic device that is most often used in literary texts. They can also be found in humorous utterances which rely on a type of complicity between the speakers (just like for some euphemisms, where the euphemistic process is obvious both for the sender and the receiver). But the advertising and political contexts also provide a breeding ground for terms which seem opposite at first sight but which are very seriously juxtaposed. This is the case, for instance, of the term *negative growth* which is defined in *Wiktionary* as follows:

> (*economics*) The opposite of economic growth; economic decline.

DEFINITION 28: *negative growth* [WIKT]

This definition may seem puzzling ; why should one talk about *negative growth* and not about *economic decline* or *economic decrease*? The definition is both antonymous and synonymous: the opposite of *economic growth* is *economic decline*, and replacing *economic* by *negative* to refer to an opposite is far from being natural. The term is not found in UD ; if it was, the definition would most probably include an analysis of its motivation and would show the underlying euphemizing process ("motivated for the speaker but not the hearer").

If we turn to another field, the term *SUV* (or *sport-utility vehicle*) can seem bewildering to someone who is not a motor industry connoisseur. First, juxtaposing *sport* and *utility* can be considered as an oxymoron (and French speakers, by whom *SUV* has been borrowed, usually have no idea what the abbreviation stands for). Second, reading the following definitions in the OED (29) or the MMD (30) does not really help understand what makes a *SUV* different from

> *n.* orig. and chiefly *North American* a four-wheel drive motor vehicle that can be used for recreational off-road driving (abbreviated *SUV*).

DEFINITION 29: *sport utility vehicle* [OED, under sport]

> sport utility vehicle: a four-wheel drive

DEFINITION 30: *SUV* [MMD]

> A passenger vehicle which combines the towing capacity of a pickup truck with the passenger-carrying space of a minivan or station wagon together with on- or off-road ability.

DEFINITION 31: *SUV* [WIKT]

a traditional *four-wheel drive vehicle*. The definition 31 found in *Wiktionary* includes more details but does not provide any answer to the question asked above, and does explain in what ways SUVs are related to sports. The definitions 32 and 33 found in UD – whose style is very typical of this dictionary – provides the reader with a few useful clues as to why there might be some misunderstandings.

> Sport Utility Vehicle. Neither a sport vehicle nor a utility vehicle. A whack, fakeass (and successful) attempt by the motor vehicle industry to lure in overpaid middle class workaholic moms who think that they need a 3 ton vehicle to carry their stupid kids to soccer practice.

DEFINITION 32: *SUV* [UD, def. #1/80, 2003]

> Gas-guzzling motor vehicle designed for off-road driving while only 5% of SUV owners ever go off road

DEFINITION 33: *SUV* [UD, def. #3/80, 2003]

According to the OED, the first occurrence of *sport-utility vehicle* goes back to 1969. The following evidence (def. 34) was produced some 30 years later, and highlights the polluting feature mentioned in (33):

> 2000 *Star-Ledger* (Newark, New Jersey) 4 Jan. 12/3 President Clinton announces a plan to make sport utility vehicles meet the same emissions standards as cars to reduce air pollution.

DEFINITION 34: *sport utility vehicle* (example) [OED]

The definition in *Wiktionnaire* (35) is very similar to the one found in *Wiktionary*, but there is also a very telling example (36), taken from a weekly magazine, about a *vrai faux 4x4*.

> Voiture bicorps initialement tout-terrain souvent à roues motrices dont l'espace utile commun aux passagers et aux bagages est modulable

<div align="center">DEFINITION 35: SUV [WIKTFR]</div>

> *"Comme on doit regretter, à Boulogne-Billancourt, de ne pas avoir disposé plus tôt d'un **SUV** comme le Renault Captur dans la gamme, tant ce vrai faux 4x4 fait un carton !"* – (Cristian DAVID, « Renault Captur : le SUV version sourire », *L'Express*, 26/02/2014)

<div align="center">DEFINITION 36: SUV (example) [WIKTFR]</div>

What the expression *vrai faux 4x4* (literally *"true-fake 4x4"*) actually means is not clarified and relies on assumed mutual understanding with the reader. This term somehow relates to the more explicit definition found in (32): "neither sport nor utility". We can notice that the factual definition where no point of view is clearly expressed is supplemented by an example that more or less implicitly refers to an additional semantic feature or connotation. This is particularly striking when we compare example (36) with example (37) that was substituted for it on Sept 22, 2017:

> *Nés pour les loisirs et la famille, les **SUV** partent aussi à l'assaut des villes grâce à des modèles de poche... Le succès des **SUV** citadins vient de leur design [...]* (*L'Argus*, consulté le 22 septembre 2017)

<div align="center">DEFINITION 37: SUV (new example) [WIKTFR]</div>

The comment that was left by the Internet user who initiated this substitution mentions he/she found it more interesting because SUV characteristics were mentioned and there was no advertising, unlike example (36) that referred to a specific car brand (*Renault Captur*). Ironically, the new example presents SUVs in a favorable light, which in itself could be seen as denoting a point of view.

Béjoint [2015] made the following comment about general dictionaries:

> Modern DGUs want to be scientific, and they succeed in being impersonal; some would say that they are boring.

This comment could very well apply to *Wiktionary*: by sticking to an editorial policy that will not depart from the neutrality principle, not only could this

dictionary be considered "boring" by some, it also fails to provide some essential defining elements. On the contrary, UD can be criticized for its lack of seriousness but the analysis of the motivation of some neologisms that is only found in this dictionary can be a useful supplement to the information found in other dictionaries.

3 Which Experts for the Lexicographic Description of New Specialized Terms and Culture?

Sablayrolles [2006] uses the term "paléologismes" to refer to lexical items which are not used any longer and which have been coined again by speakers who were not aware of their previous existence, in order to designate a new referent. Conversely, a concept or an object from the real world that already exists may be viewed as new and may then be assigned a new designation. In that case, one may wonder if there already was a lexical unit to refer to that reality and, if so, if it perfectly matches the neologism semantically speaking. For instance, spreading rumors or erroneous information is far from being a new practice. But owing to the volume of information circulating through mass media and social networks, to the pace at which it circulates and to its potential consequence (e.g. on the results of an election), speakers have felt the need to coin lexical items such as *fake news* and *post-truth* (era). In the same way, but in a different field, bikes without a freewheel system are nothing but new (they were actually initially conceived as such), but the trendy picture of the *New York courier* has led to massive sales of *fixies* to numerous *hipsters*. The French language usually resorts to equivalents created to be used in place of Anglicisms: the term *(vélo à) pignon fixe* was thus coined when the object it designates was only used by the precursors ; when those bikes became "mainstream" – *vélo à pignon fixe* became the term officially recommended in December 2013 –, the term *fixie* was then borrowed. However, this type of bike had been known for a long time as *fixed-gear bike* by professional cyclers, who used it for training or for bike races in velodromes. Then, when hipsters realized that freewheels actually were useful, this device was added to fixies, while the single-speed principle was maintained, for the sake of simplicity and elegance. *"Fixies-with-a-newly-freed-wheel"* thus had to be renamed: they became *single-gear / single-speed bikes* (*vélos single-speed* in French) – something that did not just spring into existence but that had existed for years before becoming all the rage, before that neologism was coined (and well before derailleurs were invented, for that matter). Through this example, we can see that one of the main issues that need to be addressed when begin-

ning the analysis of neology is the following: is the innovation linguistic or conceptual? Sections 3.1 and 3.2 provide two examples of neologisms whose referents did not exactly exist as such before the lexical units were coined, but whose history is essential to know, together with the culture they belong to, in order to be able to answer the question and to properly define them. The case of *hacker*, which is analyzed in section 3.3, despite not being a neologism *per se* today, is a good example of a term with a low level of technicity but a high level of polysemy which has been a challenge for terminologists.

3.1   Graphic novel, Comics, Comic book and Bande dessinée

The readers of American comic books, just like the readers of French or Belgian *bandes dessinées* (*BDs*) may wonder why, for the last couple of years (or decades for members of the *avant-garde*), some of these books have been called *graphic novels* (*romans graphiques* in French). The three definitions that follow each include a different hypernym or genus proximum: *story*, *comic book* and *novel*.

> graphic novel, a full-length (esp. science fiction or fantasy) story published as a book in comic-strip format.

DEFINITION 38: *graphic novel* (under *graphic*)
[OED, Draft additions 1993]

> a novel that uses drawings to tell the story

DEFINITION 39: *graphic novel* [MMD]

> (*comics*) A comic book containing a single full non-episodic story.

DEFINITION 40: *graphic novel* [WIKT]

The OED defines graphic novel as a full story, specifies its physical format (*book*) and the way the drawings are organized (*comic-strip*). According to the MMD, it is first and foremost a novel, and the drawings are only used for narration (we cannot tell whether the drawings are used alongside with the text or instead of the text). For *Wiktionary*, a graphic novel is a comic book whose specificity lies in its completeness with no episodes. The contrastive "non-episodic" feature is confirmed when one looks at the definition (40) *Wiktionary* provides for *comic book* which includes the term *serialized*.

> (*comics*) A book or magazine that uses sequences of drawings to tell a story or series of stories, primarily in serialized form, usually fiction.
>
> **Usage notes**
> Although the name comic book implies that humor is involved, not all comic books are funny. Oftentimes people associate comic books with adventure stories involving superheroes.

DEFINITION 41: *comic book* [WIKT]

The usage note included in *Wiktionary* points to the optional presence of humor in comic books, which is often wrongly perceived as being an inherent quality by the general public. As for the French reader, looking up the term *roman graphique* in *Wiktionnaire* (def. 42) could leave him/her feeling dissatisfied.

> Bande dessinée ambitieuse s'adressant plus spécialement aux adultes.

DEFINITION 42: *roman graphique* [WIKTFR]

According to that dictionary, the main target audience of *romans graphiques* is adults. The adjective *ambitieuse* ("ambitious"), which is rather vague, might be a tentative way of defining it as lacking the humoristic feature which is rightly or wrongly (see definition 41) associated with *comic books* (and maybe with *bandes dessinées*). The definition 43, taken from UD, seems to imply that *comic books* are books for adults who will not admit they are somehow immature. This definition even seems to suggest that the lexical unit graphic novel does not refer to a new concept but is only a new designation for comic book that erases its juvenile connotation. Another contributor (def. 44) thinks this widespread belief is incorrect, and that the new term actually refers to a particular format.

To conclude, what makes a graphic novel specific: its medium, its format, its completeness or its target audience? As far as the last distinctive feature is

> "Graphic Novel" is basically a comic book. Graphic Novel is the term mainly used by adults to make comics seem less childish

DEFINITION 43: *graphic novel* [UD, #2/4, 2005]

> what hipsters, idiots, and the media use to attempt to sound high
> brow about reading/discussing comic books, but in reality is a term
> describing a specific format of comic book. that *(sic)* form being
> an over-sized self-contained single original story (as opposed to a
> typical single issue or collection of old single issues).

DEFINITION 44: *graphic novel* [UD, #3/4, 2012]

concerned, one may still wonder if it is aimed *especially* at adults (see definition 42) or if the term is nothing but a designation with a favorable connotation used to make comic books sound more legitimate. A quick look at the entry for *roman graphique* in the French *Wikipedia* – on which definition 42 is based and from which the contributors borrowed the adjective "ambitieuse" – reveals that it is rather aimed at adults, but that the expression is also used for long *bandes dessinées* aimed at young people. Adding up the defining elements found in *Wiktionary* and *Wikipedia* ("single full non-episodic story" and "ambitieux", "public adulte") turns out to be tricky. Since *graphic novel* literally translates as *roman graphique*, one may think the concepts are equivalent. But definitions 40 and 42 include *comic book* and *bande dessinée* as *genus proximum*, and these two terms are not translational equivalents. Why is it so hard to identify what these books encompass might stem from the fact that there is no real equivalence in the cultural spheres they belong to – a *bande dessinée* is no *comic book*, a *comic book* is no *manga*, etc. – or rather, the spheres they *originate* from, since all these types of books are now translated and exported. Leaving linguistics aside to turn to cultural studies proves enlightening. J.-P. Gabilliet, whose main research interest is popular culture and comics, especially in the North American context, draws a list of differences between *BDs*, *mangas* and *comic books* [Gabilliet, 2009] and points out that *BDs*, which he calls a "media culturellement subalterne" ("a culturally subordinate type of medium") tends to be viewed through exogenous characteristics, unlike *mangas* and *comics*, whose designation seems to imply that their contents is not very serious:

> Si l'expression française « bande dessinée » [...] a le mérite de
> désigner la forme du moyen d'expression, fût-ce de manière restrictive, ses traductions anglaises (comics) et japonaises (manga,
> littéralement « images dérisoires ») mettent l'accent sur la nature
> légère des contenus, donc sur un genre qui serait l'apanage entier
> et exclusif du moyen d'expression – et pourtant, l'humour n'est
> depuis fort longtemps qu'un des multiples registres dans lesquels
> peut s'exercer l'art des auteurs de bandes dessinées.

According to Gabilliet, the difficulty in defining *bande dessinée* (and thus *roman graphique*) in terms of target audience may be accounted for in France and Belgium by the will to bring together several generations in the decades that followed the Second World War. Eliminating the juvenile connotation seems to have gone hand in hand with substituting books to magazines. In Europe, the "respectabilization" of the BD very gradually took place, starting in the 1960s, through this shift in medium. The North-American market is thought to have undergone Europeanization since the beginning of the 21st century when the paperback magazine substituted for the album format imported from Europe, which was believed to be more serious. The use of the term *graphic novel* could be considered as the ultimate stage of respectabilization. According to Gabilliet [2005], it even was Richard Kyle's objective when he coined that neologism back in 1964 in order to put an end to the negative connotations of the term *comics*. The commercial industry, Gabilliet writes, got hold of the term to make it a generic designation that would make the object just as culturally respectable as literary novels and extend this respectability to various types of publications. Gabilliet concludes his paper by saying that the term *graphic novel* does not refer to the same thing for connoisseurs or for the book market, the latter distinguishing three different subcategories (to which we have added boldface):

> **Pour critiques et esthètes,** le « graphic novel », à l'instar du roman, se conçoit dans la perspective littéraire d'une œuvre produite par un auteur manifestant une démarche créative pleine, entière et autonome. Mais **pour le marché du livre aux États-Unis,** la catégorie **« graphic novel » recouvre en fait trois réalités distinctes : 1)** les recueils de bandes dessinées de presse, **2)** les recueils de bandes prépubliées dans des comic books grand public (mettant en scène le plus souvent des super-héros ou des types de personnages apparentés), **3)** les albums contenant des histoires complètes prépubliées ou non, sans rapport étroit avec les genres grand public et constituant par rapport à ceux-ci une production indépendante.

It is rather clear then that a lexicographical description (or several, for that matter) cannot thoroughly render the various subtleties found in that field or the various perceptions of connoisseurs and professionals. Even the most tenacious lexicographer will not be able to take the time, for what seems at first sight to be such a simple entry, to lead a comparative analysis of the international history of stories narrated through a series of drawings.

3.2   Hackathon

Hackathon is a new OED entry that was published in March 2017, together with *hackfest*, *hackability* and *hackable*. It is missing from the other English dictionaries under study, except the – crowdsourced – MMOD, but can be found in the same year edition of the French PR, which is surprisingly fast. Even though this term belongs to a specialized field (computer science), its low technicity level and the transparent word-formation process that was used to coin it (a blend based on *hack* and *marathon*) make it understandable at first sight by any layman. Writing a lexicographical definition for that term can thus seem easy and resorting to an expert may seem unnecessary. The five definitions below (45 to 49), which are taken from French and English professional and crowdsourced dictionaries, seem to tally. The hypernym used at the beginning of the definition is the same in all five cases: *event / événement* in French. A hackathon thus is an event whose aim is to program, or develop software. As a consequence, a hackathon gathers participants with programming skills: *computer programmers* (OED and MMOD), *software developers* and *hackers* (OED), *programmeurs* (WIKTFR), and *spécialistes* (PR). The PR's lack of specificity leads the reader to assume that *spécialistes* actually means *computer programming specialists*. As to *Wiktionary*, it provides additional information by saying that a hackathon is an event during which "programmers and *others*" meet. The PR and the OED disagree on the length of the event: for the former, it lasts several days, while for the latter, it typically (but not necessarily) lasts several days.

> 2. A collaborative computer-programming event, typically lasting several days and involving computer programmers, software developers, hackers, etc.; a hackfest.

DEFINITION 45: *hackathon* [OED]

> An event where programmers and others meet for collaborative software development.

DEFINITION 46: *hackathon* [WIKT]

> an event when computer programmers meet to create computer programs and software applications

DEFINITION 47: *hackathon* [MMOD]

> Événement où des programmeurs se rencontrent pour collaborer au développement de logiciels.

DEFINITION 48: *hackathon* [WIKTFR]

> ANGLIC. Évènement au cours duquel des spécialistes se réunissent durant plusieurs jours autour d'un projet collaboratif de programmation informatique ou, PAR EXT., de création numérique.

DEFINITION 49: *hackathon* [PR]

Let us now examine a terminological definition for *hackathon* (def. 50).

> Rassemblement de programmeurs qui, pendant un ou plusieurs jours, compétitionnent en équipe dans le but de développer des programmes sur un thème ou pour un événement déterminé, à partir de données qui sont mises à leur disposition.

DEFINITION 50: *hackathon* [GDT]

Two new elements are found in the GDT: there is a competitive side to the event ("les programmeurs *compétitionnent*"), and the participants have data at their disposal. Even though the team aspect can relate to the collaborative development found in the definitions mentioned above, it is only in the GDT that the idea of competition is recorded. According to that dictionary, this event lasts one to several days.

Let us now examine the definitions found in UD (51) and JargonF (52).

Even though these definitions are informative, they are not necessarily understandable by readers who are not familiar with the culture of software development. What the UD contributor criticizes here is the fact that some companies organize internal hackathons, sometimes at night or over the weekend,

> Having company employees come in and work all night under the guise of innovation and opportunity with little or not reward to ones self for the sole purpose of benefiting the company.
> *Please join us in our hackathon next week we will provide pizza and beer in exchange for you burning yourself out.*

DEFINITION 51: *hackathon* [UD, def. #1/2, 2011]

> [réunion] Coding party longue et sans objectif précis, sinon d'amé-
> liorer le logiciel.

DEFINITION 52: *hackathon* [JargonF]

and that those "company hackathons" are a means used by the companies to exploit its employees under false (recreational) pretenses. To fully understand this definition, the following questions need to be asked: (i) who organizes hackathons? (ii) what for (more specifically than "to develop software")? None of the descriptions listed above (i.e. 45 to 50) mentions these defining elements, either for lack of space for encyclopedic developments, or by failing to identify them. Some answers can nonetheless be found in the *Wikipedia* entry for *hackathon*:

> [...] computer programmers and others involved in software development, including graphic designers, interface designers, project managers, and others, often including subject-matter-experts, collaborate intensively on software projects.

As was the case earlier for the definition of *graphic novel*, the definition found in *Wiktionary* (46) was most probably taken from the encyclopedic entry. We finally get to know the identity of the "other" people who collaborate with computer scientists: they are "subject-matter-experts", an expression whose meaning becomes clearer as one reads the encyclopedic entry further. The various types of hackathons are listed and classified according to their goal (more precisely the type of objective of the project being developed). Hackathons can be motivated simply by "art for the sake of art" – the main objective being for instance to use a particular programming language, but can also have a militant or humanitarian goal: some hackathons are organized in order to design tools for participatory democracy, tools to improve urban traffic, or crisis management tools for natural or health disasters. This allows the reader to gain a better understanding of why "subject-matter-experts" collaborate with software developers, and to know for a fact that the "company hackathons" that are criticized in definition 51 are only one type of hackathon among many others.

The definition found in JargonF (52) is very brief and not very specific: the only new element found there is the hypernym *coding party*. Except for those who are familiar with the programming culture of the eighties and nineties, this term needs to be looked up as well in JargonF by clicking on a hypertextual link:

[réunion] Réunion dont l'objectif est, pour un groupe de personnes participant à projet, de développer en même temps et au même endroit afin de progresser plus vite et/donc de mieux se connaître, pour ensuite progresser plus vite... jusqu'à la coding party suivante.
Voir aussi hackathon, sprint, hackfest, install party.

DEFINITION 53: *coding party* [JargonF]

When reading this definition, one may wonder what makes a hackathon different from a coding party. Providing the definitions for *install party*, *copy party* and *demo party* would be necessary here to situate the principle of a hackathon in the history of "programming events or competitions", but for lack of space, we will only sum up the definitions found in JargonF and enrich them with our own knowledge of the field. *Coding party* is a generic term which has never left the circle of professional or amateur computer scientists and whose original form was *copy party*: a gathering of *crackers* who break into software in order to get an illegal copy. Those copy parties were soon followed by coding competitions (called *coding party*), which aimed to show one's programming skills, in the form of multimedia artistic creations, *demos*, which gave rise to specific gatherings called *demo parties*. There is a long history behind the notion of hackathon which cannot be documented in a general-purpose dictionary, and one cannot expect a lexicographer to know about *coding parties*, which would help define *hackathon*, all the more so as this lexical unit cannot be found in any dictionary but JargonF. However, it still is imperceptibly productive in French: Linux users associations organize *install **parties*** during which they help neophytes install the Linux operating system on their computers, and, since programming languages for children were first designed, "***coding**-goûters*" – i.e. parties in which children are *fed* while learning how to… *feed* figures into a computer – have been all the rage. If a lexicographer wonders how long a hackathon lasts, whether it is competitive or collaborative, if data are at the users' disposal or not (none of the answers to those questions being actually a necessary element for a hackathon, as may be deduced from what precedes), who can he/she turn to? All this may seem superfluous or restrictive, as definitions 46 and 48 written by amateurs themselves which mention none of this could lead us to believe. Now, if a lexicographer seeks more semantic information by using a concordancer, his/her perception will be heavily dependent on the corpus. When reading articles related to hackathons organized by *la Quadrature du Net* (an advocacy group defending the rights and freedoms of citizens on the Internet), we understand that those events aim to "provide a framework for political ac-

tivists in order to optimize, promote and keep track of their actions". [10] On the contrary, the definition found in *Le Journal du Net*, a "leading website for business executives" with a liberal bias, only confirms what was stated in the ones studied above: "un événement lors duquel des équipes [...] doivent développer un projet informatique [...] sur une période limitée, et généralement courte (une journée, une nuit, un week-end)". [11] Regarding the participants' motivation, one can read that "les développeurs y voient un moyen de se tester, sous pression, [...] Mais, outre la gloire et la reconnaissance des pairs, il peut aussi y avoir d'autres récompenses motivantes : de l'argent, des rencontres avec des fonds d'investissement, une place dans un incubateur". This seems quite far from the idea of political *hacktivism*. The article does not deal with a particular type of hackathon but its main purpose is to provide a definition of the "hackathon phenomenon". There is no reference here to a potentially militant motivation, and, if we are to believe *Le Journal du Net*, the only existing type of hackathon is a "company hackathon". It is only towards the end of the article that we learn that hackathons "can be" organized for women, for solving water issues or to honor veterans. To conclude, let us mention that hackathons are on the way to becoming a brainstorming method that goes beyond computer science, as shown by an article (found on the website of an association which advocates implementing basic income) devoted to the "Basic Income Hackathon" which took place in Finland in March 2016 [12] and aimed to gather and promote ideas and projects related to basic income.

The various contexts in which *hackathon* is used in the sources we have just quoted corroborate what [Becker, 2015: 162] states about specialized lexicographers (which applies all the more so to lexicologists): *"Writing definitions for specialized concepts is a challenging task since the data in the corpus may present different opinions."* The corpora presented here do not exactly present different opinions but rather diverging interests. Becker adds that "in addition to consulting the corpus data, specialized lexicographers may also turn to subject field expert for guidance". But for such a borderline case (neither general nor very specialized), which expert could a lexicographer turn to? Who could consider him/herself "hackathon expert"? One last resort might be browsing through *Wikipedia*, whose reliability can be questioned. The entry dedicated to *hackathon* was created on May, 28 2005 and has been enriched since ; it

---

10. `https://wiki.laquadrature.net/Hack-a-thon1_Political_Memory_2.0`

11. Hackathon : les clés pour comprendre un phénomène qui prend de l'ampleur, 23/06/2014
`http://www.journaldunet.com/solutions/emploi-rh/hackathon.shtml`

12. Finlande - Le Hackathon ouvre la voie à une expérimentation du revenu de base", 18/03/2016, `www.revenudebase.info/2016/03/18/ finlande-hackathon-experimentation-revenu-de-base/`

includes the various characteristics of hackathons mentioned in the corpus extracts and definitions studied above.

3.3   Hacker

In the previous section, we mentioned that the word-formation process that led to the term *hackathon* could easily be identified since it was a (transparent) blend of *hack* and *marathon*. This, actually, could be considered as a sweeping assertion and needs further explanations. In the OED, the entry *hack* includes 7 homonyms for the noun – most of which are polysemous, and only 2 of which relate to computer science – and 6 for the verb (only one of which is related to computer science). Bringing this diversity of meanings to mind seems necessary to account for the confusing terminological and lexicographical definitions of *hack* (noun and verb) and *hacker* in English, and the even more confusing definitions of the French lexicalized borrowings *hackeur* (noun) or *hacker* (verb or spelling variant of the noun *hackeur*, pronounced differently). Nowadays, even though the way individuals picture it may vary, the term *hacker* is part and parcel of our collective imagination, and could be considered as belonging to the general language. In that case, its presence in a term bank may be questioned. But if one thinks about the field of IT security and its rules, the need for a precise definition for *hacker* can be felt: which ones of his/her acts are considered criminal?

There are several records (which are not copied here for lack of space) for the term *hacker* in the GDT and *Termium* term banks, and the most frequent defining elements are the same as those found in general purpose dictionaries: presence of malicious intent or not, level of competence (is a hacker skilled or just opportunistic?), use of a systematic or of a random empirical method (is a hacker smart or tedious?). According to the GDT, a *hacker-bidouilleur* acts randomly. Term banks include fuzzy statements or even internal contradictions, as is the case in *Termium Plus*, in which one learns that a hacker "utilise ses connaissances techniques étendues" and is "un programmeur créatif", but also that the confusion about that term leads one to believe that such a person must be gifted, which is not necessarily the case since any patient, tenacious person with enough information can easily break into a computer system. [13]

---

13. « la confusion laisse à penser qu'un tel individu doit être talentueux et brillant, ce qui n'est pas nécessairement vrai puisqu'il suffit de bénéficier de renseignements suffisants et de faire preuve de patience et de persévérance pour pénétrer un système ». *Termium Plus, hacker* (fiche 1).

Here is how the verb *hacker* is defined in the French general-purpose dictionary PR:

> ANGLIC. Pirater (un système, un compte informatique) par jeu, goût du défi sans intention de nuire. *Le site du ministère a été hacké.*

DEFINITION 54: *hacker* [PR]

The definition states that people can *hack* with no malicious intent ("sans intention de nuire"), but there is a blatant contradiction in the example provided: whether classified information was stolen from the ministry's website or modified, or whether the website was crashed obviously results from a malicious behavior. Rogers [2013], who studied the thorny issue of defining *hacker*, asserts that a distinction needs to be made between *crackers* and *crashers*. Classifying hackers into subtypes has revealed that "many other sub-categories of hacker with various labels and even a so-called hacker definition controversy about the "true" meaning of this core term either as benign (as in the original sense of a knowledgeable computer enthusiast) or malign (as in Data Security and in the media)." In French, the etymological information provided by the PR for the headwords *hackeur* (def. 55) and *hackathon* (def. 56) just add to the overall sense of confusion: they both refer to the same English etymon, the verb *hack*, but with two very different translations.

> 1984 ◊ mot anglais, probablt de *to hack* au sens argotique de « perdre son temps »

DEFINITION 55: *hackeur* (etymology) [PR]

> 2010 ◊ mot-valise anglais, de *(to) hack* « être passionné d'informatique » et (mar)athon

DEFINITION 56: *hackathon* (etymology) [PR]

Even though a *hacker* necessarily is a computer science fanatic (a "passionné d'informatique"), *hack* cannot translate as *being* a computer science fanatic as suggested in (56). Moreover, the meaning found in (55) (to waste one's time) could correspond to the phrasal verb *hack **around***, but not *hack*, and certainly cannot be the best explanation for what can be paraphrased in French as *marathon de programmation*. Finally, it is unfortunate that the noun *hack* is nowhere to be found with its computer science meaning – be it in the PR

or in the term banks (in which the only meanings recorded are "cheval" and "taillade"). A possible translation in French could be *bidouille*, with which either the positive connotation of *astuce* (= *clever technical trick*) or the negative connotation of *rustine* (= *kludge*) can be associated, as clearly stated in JargonF's entry for *bidouille* (def. 57).

> 1. « La bidouille ». Programmation pas du tout robuste, crado, pas documentée, mais qui fonctionne, parfois.
> 2. « Une bidouille ». Petit bout de programme, petite réalisation technique, permettant d'obtenir des résultats avec peu de moyens.

DEFINITION 57: *bidouille* [JargonF]

Lastly, let us now examine the two definitions for *hacker* found in UD, which successfully gather *all* the semantic features and connotations usually associated with the term.

> An individual capable of solving complex non-intuitive problems in a seemingly intuitive manner. The processes and techniques used are not necessarily methodical to the observer, but yet achieve results significantly and consistently faster than known experience would predict. A hacker is not defined in terms of intention or purpose, but rather by the talented single-mindedness of method. Hackers are not limited to computer hacking.
> Commentary: The movies "Tron" (1982) and "War Games" (1983) significantly influenced the common use of "hacker" (by non-hackers) in reference to computer hackers. [...]

DEFINITION 58: *hacker* [UD, def. #5/84, 2006]

Definition 58 first reminds the users that, unlike what was captured by the collective imagination based on the movie *War Games*, the term does not necessarily relate to the field of computer science. It then explains that even though the method used by the hacker does not seem methodological, it might only be a matter of perception on the observer's part. The last clarification introduced by (58) is that the term hacker should not be defined based on his/her goal, which puts an end to the debate over his/her potentially malicious intents. Definition 59 provides an account of the evolution of the meaning of the term over the last decades, presents the *casual / black hat / white hat* classification (in which objectives can be mentioned, but only to show how

diverse they can be) and alludes to the wrong, obsolete, usage of the term among computer illiterates.

> Back in the 70's, hacker was a term given to those fortunate enough to know how to code. During the 80's and 90's it was a term meant for those who worked their way through systems, without approval. Now, its meaning has been completely replaced by cracker, and hacker means nothing more than any idiot that can decipher a small page of HTML. Those who are computer illiterate still widely use the word in its 80s/90s sense. [...] Of the 80s/90s version, there were three primary denominations:
> The casual Hacker-hacks to learn information for his own curiosity.
> The White Hat Hacker-hunts down and destroys malicious code.
> The Black Hat Hacker-designs and releases malicious code;gathers dangerous information;brings down sensative systems (*sic*) [...]

DEFINITION 59: *hacker* [UD, def. #6/84, 2005]

Hacker and *hackathon* each have a story of their own. *Hacker* might be in the process of having its connotation changed, just like the one analyzed by Galinsky et al. [2003] for *geek*. Such a change results from "in-group planning" but is also a case of spontaneous neology. On the one hand, geeks which have started calling themselves *geeks* prevent other people from using that term with a negative connotation (another example of reappropriation of a stigmatizing label is the word *gay*). On the other hand, the "dot-com revolution" and the picture of the Silicon Valley billionaires closely associated with it have made "the link between computer aptitude and economic success" rather obvious. Being able to climb up the social ladder – be it symbolically or for real – now is one essential semantic feature, but the "social skills deficiency" geeks are traditionally characterized by should not be left apart. The teenage genius from *War Games* has gained a professional status. The respectable reputation of hackathons (or at least of "company hackathons") could play a part in improving the connotations of the term *hacker* further. Analyzing the partial semantic or connotational shift of this term that has made its way into the general language at the same time as computer science became part of our everyday life could be an opportunity to look back at the original meaning of the term and how it has evolved.

Conclusion

The main objective of this paper was to show how crowdsourced dictionaries complement professional dictionaries, and to which extent, through the filter of neology. The emphasis has been in particular on some kinds of neological phenomena which are especially convincingly treated in crowdsourced dictionaries.

Another objective of this paper was to shed light on the potential of amateur dictionaries for the study of neology – not by casting doubt on the usefulness of corpus linguistics or questioning the lexicographical descriptions of professional dictionaries, but rather by proving how this particular type of dictionary can be viewed as complementary tools in many respects.

As far as neologism detection is concerned, *Wiktionary* could undermine the consensual idea that a neologism is a lexical unit which has not been recorded in a dictionary yet: due to its editorial policy and the high number of contributors on the lookout, its nomenclature is nonstandard, and new terms are granted an entry in record time. Instead of being used as a traditional corpus of exclusion, Wiktionary could then act as the opposite end of the spectrum, i.e. as a tool for neology watch.

As far as the description of the semantic features of a neological unit is concerned, since the boundaries between the general language and specialized languages, between general culture and subcultures are often blurred, it may rely both on a lexicographical and a terminological approach. The type of expertise required is not only technical or scientific: knowledge of the very culture of the field is also needed. Owing to the wide range of professional and cultural backgrounds found among the Internet users and contributors, amateur dictionaries cover a vast array of specialized fields. This is the case for *Wiktionary*, but even more so for specialized amateur dictionaries such as JargonF. Not only does that dictionary provide precise definitions of the most technical terms, it also gives information on what Charnock [1999] calls "vague and indeterminate terms", which are used to designate "fundamentally fuzzy concepts". Providing new insights into that "conceptual fuzziness" sometimes requires introducing a point of view – such as the short comments found in JargonF –, which is not standard practice in more conventional resources.

Finally, amateur dictionaries prove useful regarding (spontaneous or planned)

semantic neology, where the new meaning of a lexical unit is contrasted with its original meaning. The latter can best be found in traditional dictionaries. However, when a speaker notices a striking difference between what he/she has just read or heard and what is recorded in a traditional dictionary, he/she may turn to a crowdsourced dictionary where he/she will find examples of prescriptive metadiscourse written by contributors who criticize the "wrong" use of a term, which they blame on the (other) speakers' lack of linguistic knowledge. When using *Urban Dictionary*, they can also have access to "deviant" meanings due to lexical manipulations, whose motivation is analyzed by the contributors. Could this unique dictionary, which may be labelled a "sense-checking tool", deter us from entering a "post-semantics era"?

References

Ager, D. (2001). *Motivation in Language Planning and Language Policy*. Multilingual Matters.

Atkins, S. and Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford University Press, Oxford.

Becker, H. (2015). Scientific and Technical Dictionaries; Coverage of Scientific and Technical Terms in General Dictionaries. In Durkin, P., editor, *The Oxford Handbook of Lexicography*. Oxford University Press, Oxford.

Béjoint, H. (1998). Scientific and technical words in general dictionaries. *International Journal of Lexicography*, 1(4):354–368.

Béjoint, H. (2015). Dictionaries for General Users: History and Development; Current Issues. In Durkin, P., editor, *The Oxford Handbook of Lexicography*. Oxford University Press, Oxford.

Bowker, L. (2003). Specialized lexicography and specialized dictionaries. In Van Sterkenburg, P., editor, *A Practical Guide to Lexicography*. John Benjamins, Amsterdam/Philadelphia.

Calabrese, L. (2018). Faut-il dire *migrant* ou *réfugié* ? Débat lexico-sémantique autour d'un problème public. *Langages*, 210(2):105–124.

Campbell, L. (1998). *Historical Linguistics. An Introduction*. MIT Press, Cambridge.

Čermák, F. (2003). Source materials for dictionaries. In Van Sterkenburg, P., editor, *A Practical Guide to Lexicography*. John Benjamins, Amsterdam/Philadelphia.

Charnock, R. (1999). Les langues de spécialité et le langage technique : considérations didactiques. *ASp*, 23-26:281–302.

Corbin, P. (1998). La lexicographie française est-elle en panne ? In *Cicle de Conferències 96-97, Lèxic, corpus i diccionaris*, pages 83–112, Barcelona.

Corbin, P. (2008). Quel avenir pour la lexicographie française ? In *Congrès Mondial de Linguistique Française*, pages 1227–1250, Paris, France.

Crespo-Fernández, E. (2014). Euphemism and political discourse in the british regional press. *Brno studies in English*, 40(1):5–26.

Damaso, J. and Cotter, C. (2007). UrbanDictionary.com. Online dictionaries as emerging archives of contemporary usage and collaborative lexicography. *English Today*, 23(2):19–26.

Díaz Hormingo, M. T. (2012). Lexical creation and euphemism: Regarding the distinction *Denominative or Referential Neology vs. Stylistic or Expressive Neology*. *Lexis*, 7:107–120.

Dubois, J. and Dubois, C. (1971). *Introduction à la lexicographie : le dictionnaire*. Librairie Larousse, Paris.

Falk, I., Bernhard, D., and Gérard, C. (2014). From Non Word to New Word: Automatically Identifying Neologisms in French Newspapers. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC 2014)*, pages 4337–4344, Reykjavik.

Farina, A. (2005). Lexicographie et discrimination. Corso on line - Introduzione agli studi di genere.

Farina, D. M. T. C. (2016). Lexical change in times of upheaval and war - and the dictionary. In *Proceedings of the 17th EURALEX International Congress*, pages 767–776, Tbilisi, Georgia.

Gabilliet, J.-P. (2005). Du comic book au graphic novel : l'européanisation de la bande dessinée américaine. *Online Magazine of the Visual Narrative*, 12:35–40.

Gabilliet, J.-P. (2009). BD, mangas et comics : différences et influences. *La Revue*, 54(2):35–40.

Galinsky, A. D., Hugenberg, K., Groom, C., and Bodenhausen, G. (2003). The reappropriation of stigmatazing labels: Implications for social identity. *Identity Issues in Groups (Research on Managing Groups and Teams)*, 5:221–256.

Gao, Y. (2012). Online english dictionaries: Friend or foe. In *Proceedings of the 15th EURALEX International Congress*, pages 422–433, Oslo,Norway.

Girardin, C. (1979). Contenu, usage social et interdits dans le dictionnaire. *Langue française*, 43:84–99.

Hathout, N. and Sajous, F. (2016). Wiktionnaire's Wikicode GLAWIfied: a Workable French Machine-Readable Dictionary. In *Proceedings of the tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia.

Hathout, N., Sajous, F., and Calderone, B. (2014). GLÀFF, a Large Versatile French Lexicon. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland.

Krieg-Planque, A. (2007). Souligner l'euphémisme : opération savante ou acte d'engagement ? Analyse du « jugement d'euphémisation » dans le discours politique. *Semen*, 17.

Krieg-Planque, A. (2012). Dictionnaires, glossaires et lexiques militants : pratiques profanes de la critique du langage politique. In Aubry, L. and Turpin, B., editors, *Victor Klemperer. Repenser le langage totalitaire*, pages 299–313. CNRS Editions, Paris.

Krieg-Planque, A. (2015). Construire et déconstruire l'autorité en discours. le figement discursif et sa subversion. *Mots. Les langages du politique*, 107:115–131.

Le Draoulec, A., Péry-Woodley, M.-P., and Rebeyrolle, J. (2014). Glissements progressifs de "sémantique". *Le Discours et la Langue*, 1(6):109–126.

Lecolle, M. (2012). Sentiment de la langue, sentiment du discours : changement du lexique, phraséologie émergente et "air du temps". *Diachroniques*, 2:59–80.

Martinez, C. (2009). Une base de données des entrées et sorties dans la nomenclature d'un corpus de dictionnaires : présentation et exploitation. *Études de linguistique appliquée*, 156:499–509.

Meyer, I. and Mackintosh, K. (2000). L'étirement du sens terminologique : aperçu du phénomène de la déterminologisation. In Béjoint, H. and Thoiron, P., editors, *Le sens en terminologie*, pages 198–217. Presses Universitaires de Lyon.

Mortureux, M.-F. (2011). La néologie lexicale : de l'impasse à l'ouverture. *Langages*, 183(2):11–24.

Nida, E. (1995). Lexical cosmetics. In Kachru, B. B. and Kahane, H., editors, *Cultures, Ideologies, and the Dictionary. Studies in Honor of Ladislav Zgusta*, pages 69–72. De Gruyter.

Pruvost, J. (2005). Quelques concepts lexicographiques opératoires à promouvoir au seuil du XXIe siècle. *Études de linguistique appliquée*, 137(1):7–37.

Pruvost, J. (2006). *Les dictionnaires français ; outils d'une langue et d'une culture*. Ophrys, Paris.

Rey, A. (1995). Du discours au discours par l'usage : pour une problématique de l'exemple. *Langue française*, 106:95–120.

Rey-Debove, J. (1971). *Étude linguistique et sémiotique des dictionnaires français contemporains*. Mouton, Paris – La Haye.

Rogers, M. (2013). What is a 'domain' and is this a useful question? *ASp*, 64:5–16.

Sablayrolles, J.-F. (2006). La néologie aujourd'hui. In Gruaz, C., editor, *À la recherche du mot : De la langue au discours*, pages 141–157. Lambert-Lucas.

Sablayrolles, J.-F. (2008). Néologie et dictionnaire(s) comme corpus d'exclusion. In Sablayrolles, J.-F., editor, *Néologie et terminologie dans les dictionnaires*, pages 19–36. Champion, Paris.

Sajous, F. and Hathout, N. (2015). GLAWI, a free XML-encoded Machine-Readable Dictionary built from the French Wiktionary. In *Proceedings of the of the eLex 2015 conference*, pages 405–426, Herstmonceux, England.

Sajous, F. and Hathout, N. (2017). Informativité, neutralité et point de vue dans une offre dictionnairique hétérogène : vers une complémentarité ? *Revue française de linguistique appliquée*, XXII(1):27–39.

Sajous, F., Hathout, N., and Calderone, B. (2014). Ne jetons pas le Wiktionnaire avec l'oripeau du Web ! Études et réalisations fondées sur le dictionnaire collaboratif. In *Actes du 4e Congrès Mondial de Linguistique Française (CMLF 2014)*, pages 663–680, Berlin.

Sajous, F., Josselin-Leray, A., and Hathout, N. (2018). Définir la néologie terminologique dans les dictionnaires généraux : le domaine de l'informatique analysé par « les foules » et par les professionnels... de la lexicographie. In *4ème Congrès international de néologie des langues romanes (CINEO 2018)*, Lyon.

Sommant, M. (2000). Innovation lexicale : sources des néologismes, normalisation et intégration dans les nomenclatures des dictionnaires de langue français. In Béjoint, H. and Thoiron, P., editors, *Le Sens en terminologie*, pages 247–260. Presses Universitaires de Lyon.

Dictionaries

GDT	*Le Grand Dictionnaire Terminologique* [`www.granddictionnaire.com`]
JargonF	*Le Jargon Français, dictionnaire d'informatique francophone* [`jargonf.org`]
LDELC	*Longman Dictionary of English Language and Culture, 2nd edition (1998)*
MMD	*Macmillan Dictionary* [`www.macmillandictionary.com`]
MMOD	*Macmillan crowdsourced Open Dictionary* (part of MMD)
OED	*Oxford English Dictionary* (paid access) [`www.oed.com`]
PR	*Petit Robert de la langue française* (paid access) [`pr.bvdep.com`]
TERM	*Termium Plus* (The Government of Canada's terminology and linguistic data bank) [`btb.termiumplus.gc.ca`]
UD	*Urban Dictionary* [`www.urbandictionary.com`]
WIKT	*English Wiktionary* [`en.wiktionary.org`]
WIKTFR	*Wiktionnaire* (the French language edition of *Wiktionary*) [`fr.wiktionary.org`]