Bulletin of the *Transilvania* University of Braşov
Series IV: Philology and Cultural Studies • Vol. 8 (57) No. 2 – 2015

The future of dictionaries and term bases

Attila IMRE[1]

The present article focuses on similarities and differences between printed and online dictionaries, as well as term bases. Starting from a central idea that the revolution of translation resulted in the development of computer-assisted translation tools, we argue that the quality (reliability) of a term base is a basic requirement for a professional translator, who has to take advantage from all possible online and offline resources. We offer examples of possible problems in both dictionaries and term bases, then two terms are compared in matches from printed dictionaries and an online dictionary / term base. The conclusions try to foreshadow the possible future of these resources based on present developments in the age of globalization and localization.

Key-words: *dictionary, term base, law, guidelines, comparison*.

1. Introduction

The importance of printing has never been questioned. In fact, it was considered even by Francis Bacon as one of the three inventions (together with gunpowder and the compass) that changed the world (*Novum Organum,* 1620), and we truly believe that the printing of millions of books has also led the world to where we are now. Yet, a timely question is whether we still need printed books and dictionaries in a "new" age characterized by globalization, localization and the revolution of technology.

The technological developments resulted in the (r)evolution of translation as well (Imre 2013, 102–174), and e-databases are becoming more and more popular. Human translators had to face the challenges of machine translation (MT) starting from the Second World War, then – more recently – the arrival of computer-assisted translation tools (CAT), which handle translation memories (TM) and term bases (TB). Although there were many allegations that human translators would disappear and machines would take over (cf. the sci-fi movies of Hollywood), we agree with those who are reluctant to accept that this can happen in the forthcoming decades (cf. Piron[2]; Bennett and Gerber 2003, 188–189, or Kis and Mohácsi-Gorove 2008, 13). Nevertheless, it is worth taking advantage of recent technologies in case we

[1] Sapientia University Cluj-Napoca, Petru Maior University Tg.-Mureş, amireittal@gmail.com
[2] http://self.gutenberg.org/article/whebn0000019980/machine%20translation, 11. 09. 2015.

BDD-A22791 © 2015 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.216.51 (2025-12-04 20:24:10 UTC)

want to remain competitive (McKay, 2006; Samuelsson-Brown, 2010), whatever the field of interest.

We would like to compare the advantages and disadvantages of dictionaries and term bases, as basically both of them contain terms in at least two languages. A typical dictionary is "A book explaining or translating, usu. in alphabetical order, words of a language or languages, giving their pronunciation, spelling, part of speech, and etymology, or one or some of these" (Trumble, and Stevenson 2002, 673). One might add that dictionaries may be in electronic format as well (possible extensions: *.pdf, .doc, .docx, .djvu*, etc.), and an online definition already mentions this fact: "A book or electronic resource that lists the words of a language (typically in alphabetical order) and gives their meaning, or gives the equivalent words in a different language, often also providing information about pronunciation, origin, and usage."[3]

Dictionaries also offer further types of information: for instance, we can mention the grammatical category of the main entry, register and style (informal, slang, taboo, etc.), whereas TBs are not designed this way, at least for the time being. Thus we consider it important to look behind the scenes regarding their structure.

2. Possible guidelines to (legal) dictionaries and term bases

Up to the recent past, dictionaries were considered among the most authoritative sources of languages as they used to be as error-free as humanly possible. This authority is nevertheless challenged when two variants can be found in a dictionary, for the English *mental disturbance* (*tulburare mintală* and *tulburare mentală*). In most cases the *Romanian Explanatory Dictionary*[4] (Coteanu, Seche and Seche, 1996) will help, although in this case further research is needed.[5]

As the technological rush resulted in the publishing of dictionaries containing many typological, grammatical, content-related and layout-related errors (Imre, 2014a; Imre, 2014b), the most important advantage of dictionaries – reliability – seems to be shaken. On the other hand we should consider the fear of non-technical language professionals in mastering software during translations, such as TBs, combined with the time, money and energy invested in a quality database (translation memory and TB). The greatest hindrance of TBs is usually reliability, as we do not know the source(s). Thus the question is, whether we should (still) use dictionaries or find acceptable TBs. From the outset, we would opt for the combination of both, and create our own TB, as in the 21st century *quality* is the major dividing line between professionals and non-professionals.

[3] Source: http://www.oxforddictionaries.com/definition/english/dictionary, 08. 06. 2015.
[4] The online version of DEX with almost 600,000 entries is accessible here: http://www.dex.ro/, 09. 06. 2015.
[5] http://scri.ro/mental-sau-mintal-827.html/comment-page-1, 09. 06. 2015.

BDD-A22791 © 2015 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.216.51 (2025-12-04 20:24:10 UTC)

We cannot avoid *quality assurance* in translation (TB check, TM check, etc.), through which one can fight the unfortunately large and cheap possibilities of publishing low quality material in all fields, including compiling dictionaries and glossaries. Dictionaries are usually limited in size (printing costs), whereas TBs are not constrained in this respect: hundreds of thousands of entries fit into a small *Microsoft Office Excel* file (*.xlsx*), which can be easily converted into an extension compatible with CAT-tools (*.csv*). A further advantage of TBs over dictionaries is the possibility to modify them any time later; after all new entries should be added, old ones deleted or altered (e.g. new meanings added or problematic translations / meanings / explanations replaced).

The importance of these changes may be easily proved with examples. Our project to create a legal TB started with checking the published bilingual dictionaries (Romanian-English, English-Romanian) in Romania between 1999 and 2014, and we were able to find around fifteen dictionaries (with legal and / or terms belonging to economics). In one of the dictionaries, *gross indecency* is translated as *homosexualitate, pedofilie* (*homosexuality, paedophilia*), a term which was used in UK and Canada in the 1960s, but this fact is not mentioned in the dictionary. Thus, we consider that *homosexuality* cannot be the translation (or synonym) of *gross indecency* in a political correct dictionary. This is not an isolated case, as *pervers* (*perverse*) is translated as *gay* or *homosexual*, while *rasă neagră* is *Negro race*, although it should be *Afro-Americans* for the past decades. Other official terms preserve the Christian background, as *nume de botez* or *prenume* is translated as *Christian name*, whereas this term has 'non-Christian' alternatives as well: *first name* and *given name*. Thus a TB may or should contain the following terms:

Romanian	**English**
nume de botez	*Christian name*
nume de botez	*first name*
nume de botez	*given name*
numele mic	*Christian name*
numele mic	*first name*
numele mic	*given name*
prenume	*Christian name*
prenume	*first name*
prenume	*given name*

Table 1. *Terms for **first name** in Romanian and English*

Latin terms used to be more typical in 'legalese' (Imre & Barabás, 2015), contributing to the difficulty of legal terms for non-professionals who are not

BDD-A22791 © 2015 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.216.51 (2025-12-04 20:24:10 UTC)

members of the Bar. Although almost all Latin terms can be easily found on the Internet with lavish explanations (*ex nunc, nole prosequi, pro bono publico*), some of them tend to disappear. However, a good dictionary or TB should preserve them, especially when they are still in use. Thus the Romanian *pensie acordată soției pe perioada / timpul divorțului* is *temporary alimony* in English and *pendente lite* in Latin. Of course, in case the medium is changed (e.g. subtitling, dubbing), the easiest equivalent should be used, unless the point is not to understand the procedure by the client / viewers (cf. Nida's functional equivalence). Still, we can say that the emergence of McWorld and McLanguage (Barber 1992; Snell-Hornby 2006) has led or will lead to McTranslations and McDictionaries with simplified language to be more easily understood by the large public, dooming well-established Latin expressions.

The acceptance or banishment of Latin terms leads to a further issue, namely the mixture of legal terms with terms belonging to economics. As there are many forbidden, illegal activities with money, it is obvious why one cannot clearly separate these terms, although the problem is manifold. Once we accept economics within legal terms, then it comes extremely difficult to exclude other fields, such as medicine (health-related issues), biology (affected body parts, birds!), sports (misbehavior) and geography (location of an incident), as the following examples show: *rinofaringită – rhynopharyngitis, tibia – shinbone, prepeliță – quail, patinaj artistic –figure skating, Europa Occidentală – Western Europe*.

We also consider that various political, economic and other associations should be excluded as well (e.g. *FECOM, Fondul European de Colaborare Monetară – European Monetary Cooperation Fund*), as they are not strictly connected to law; similarly, countries, currencies, archaic or very rare words need not be included (*zavistie – envy, jealousy*; *zbârci – miss, wrinkle*; *zbir – brute, oppressor*).

Although some terms are substandard in written form (*fală – ring, moonlighting*), they may appear in speech, thus they may be useful for interpreters. Some less offensive terms – to a certain extent – should be included as well, for similar reasons (*nepriceput – good-for-nothing, incompetent*). At this point we should mention swear words and taboo, which is an ardent issue in subtitling. The latest trends criticize toning downs, omissions or euphemisms, arguing that it is not natural to weed out curses, blasphemies (cf. Tveit 2009, 89), but dictionaries systematically exclude them, or only neutralized versions are left. If we have in mind a bilingual dictionary or TB (Romanian and English), we should probably use the non-English standards[6]; in our case the Romanian DEX should serve as a

[6] English dictionaries contain f-words, even though the label them as *coarse slang* (Trumble and Stevenson 2002).

BDD-A22791 © 2015 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.216.51 (2025-12-04 20:24:10 UTC)

guidance. As a matter of fact, some *familiar* terms belonging to children's language (e.g. *pipi – pee-pee*) are listed in printed dictionaries (Coteanu et al. 1996, 795; Lozinschi 2008, 446) and online dictionaries[7], while others are not. As it may prove difficult which terms to preserve or not, the entire category should be excluded, as these are unlikely to be used in legal contexts during court sessions.

Yet, some might argue that there are terms 'set in stone' which are used during legal proceedings either officially or not. But this already takes us to the next section.

3. Challenging printed dictionaries

Since the appearance of machine translation (MT) there has been a constant fear of human translators that they will lose their job due to the automation of translation. But long before that, we could hear predictions about the disappearance of books in the digital age. Although MT is getting better and better, professional (human) translators still have their jobs; in fact their number is growing due to various reasons (Imre 2013, 206–229). Similarly, books are still printed, and various types of printed dictionaries (general and specialized) are mushrooming.

Hence not quantity but quality is the issue. Printed books and dictionaries should prove that they are worth considering due to reliability, although the real challenge is not the Internet, but the bridge between the immense Internet and printed dictionaries, namely *online dictionaries* and their combination with various *online term bases* and *translation memories*. In this respect we can mention the multilingual *glosbe.com*[8], or the *EUR-Lex database*[9], offering access to the European Union Law in all official European languages in the form of parallel texts.

We cannot deny that these online possibilities are surprisingly better and better. In Romanian–English and English–Romanian language combinations we can mention *hallo.ro*[10] or *ro-en.gsp.ro*[11] with remarkable results. It is obvious that online databases make use of more sources, thus it seems fair to collect as many printed dictionaries as possible in case we are to compare results.

During a POSDRU project we collected printed Romanian-English dictionaries on law between 1999 and 2014 over a period of 18 months, although the list was completed with dictionaries containing both legal and economics terms. Furthermore, legal terms were extracted from general dictionaries as well.

[7] http://www.dex.ro/pipi, 09. 06. 2015.
[8] https://glosbe.com/, which contains more than 1,000,000,000 sentences, 09. 06. 2015.
[9] http://eur-lex.europa.eu/homepage.html, 09. 06. 2015.
[10] http://hallo.ro/, 09. 06. 2015.
[11] http://ro-en.gsp.ro/, 09. 06. 2015.

BDD-A22791 © 2015 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.216.51 (2025-12-04 20:24:10 UTC)

Our first entry to be tested was the Romanian *criminal* (*criminal*) in the Romanian–English dictionaries. All in all we found 70 entries, out of which 28 occurrences were single-word terms (*criminal*) in Romanian with the following translations: ***assassin**, blameworthy, convict, **criminal**, criminally, evildoer, **felon**, **felonious**, guilty of crime, gunman, **homicidal**, **homicide**, killer, **malefactor**, man-killer, **murderer**, murderess, **murderous**, **offender**, **outlaw**, **outrageous**, penal, **perpetrator**, person who commissions a crime, principal to a crime, serious criminal, **slayer**, violent criminal*. The list contained further 42 terms with combined words in Romanian containing *criminal* (e.g. *criminal în serie – serial killer*).

Then the Romanian term *criminal* (single-word entry and compounds containing it) was searched for online, in the database of *hallo.ro* (600,000 definitions). All in all, 52 matches were found, out of which 29 occurrences were single-word terms in Romanian with the following translations: ***assassin**, crimeful, **criminal**, desperado, flagitious, **homicidal**, **homicide**, iniquitous, internecine, **malefactor**, miscreant, **murderer**, **murderous**, **offender**, **outlaw**, **outrageous**, **perpetrator**, **slayer**, tiger, wrong, sinister, con, hoodlum, lag, perp, tough, **felonious**, **felon**, wrongdoer*. Five of them are used in informal language (starting with *con*), whereas the last three are specified as belonging to legal terms. Further 23 expressions contain the Romanian *criminal* (e.g. *criminal condamnat la spânzurătoare – gallows bird*) and almost half of them belong to informal language (slang).

The results speak for themselves, offering many possible interpretations. An obvious one is that the combination of more than a dozen printed dictionaries is much more valuable than one of the 'best' (richest) online dictionary / database. However, if we were to combine more online sources, the collection could easily outnumber the results in the printed dictionaries. A further observation is that the online source contains many informal terms, being much closer to spoken English. Single-word entries are easy to compare: 14 English entries out of 28 (dictionaries) and 29 (*hallo.ro*) match, highlighted in bold, deriving from the Romanian *criminal*. As predicted, none of the English slang terms for the Romanian *criminal* could be found in the dictionaries. A more interesting fact is that dictionaries contain overwhelmingly much more expressions with *criminal*, even though the online database offers valuable terms as well (*bring a criminal to justice, mug, moll, eel,* etc.).

Yet, this is still only one side of the coin, as a different entry will lead to a completely different outcome. For instance, the Romanian *prostituată*[12] (*prostitute*) has only five occurrences in the dictionaries, whereas *hallo.ro* lists 47 (!) possible translations (all the five from the dictionaries are among them), and there are five more expressions containing the entry. 33 English translations are labelled as belonging to informal language.

[12] This is the official term listed in the Romanian Explanatory Dictionary and used in Romanian case files as well.

BDD-A22791 © 2015 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.216.51 (2025-12-04 20:24:10 UTC)

4. Conclusions

The above cases demonstrate that if we think professionally it is worth combining the printed dictionary results with the online dictionaries / term bases. However, the tendency is to transform printed dictionaries fully online, enabling further enhancements: updates (extra information, additions, deletions or corrections), and possibly with extra filtering possibilities. The larger the database, the more filtering options are welcome, detailed below.

Users may need to know subtle differences between United Kingdom or United States spelling, select archaic / obsolete forms, Latin expressions, grammatical categories, (stock) phrases, even full sentences (some printed dictionaries already contain a few full sentences deriving from phrases), or even to list all the entries beginning with a particular letter. This way the online dictionary will be suitable for the expectations of professional users in the 21st century.

The largest publishing houses already have online mono- or bilingual dictionaries (*Cambridge*[13], *Oxford*[14], *MacMillan*[15]), taking advantage of the fact that they are well-known names in the field, but the concept of going online (and mainly free of charge) signals something really important. Marshall McLuhan's *global village* not only expanded into the notorious *globalization* and *localization* pervading our McWorld, but also we are faced with the fact that *languages* (in all combinations) are turning into public domain. A clear example is the EU's term base focusing on legal terms as they recognized that the establishment is financed by the community. Another example, the Romanian Explanatory Dictionary available online at *dex.ro* is priceless, similarly to its English counterparts mentioned above, containing new and old (archaic), formal and informal (slang, *argou*), bookish (*livresc*), rare (*rar*) terms and provincialisms (*popular*).

As for dictionaries, alphabetical order may be a drawback. In the case of expressions the rule of alphabetical order does not apply, so we are faced with a limited searchability (clusters of words / expressions bunching from the main entry). The custom is to cluster entries around a headword, which seems to be subjective, at least to a certain extent. When faced with a multiple word term, the dictionary compiler has to make a choice where to place the following expressions. So *nu suferă nici o amânare* (*allows no delay*) may be under S or A (negative forms should be disregarded), but *bani fără acoperire emişi în situații de urgență* (*fiat money*) is more complicated to position under B, A, E, S, or U.

On the other hand, online dictionaries and TBs have often been criticized based on users' experience, who must have been right due to various reasons. Many online sources used to be very unreliable, but once machine translation is constantly

[13] http://dictionary.cambridge.org/, 14. 08. 2015.
[14] http://www.oed.com/, 14. 08. 2015.
[15] http://www.macmillandictionary.com/, 14. 08. 2015.

BDD-A22791 © 2015 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.216.51 (2025-12-04 20:24:10 UTC)

improving, all online sources are getting better and better since the appearance of term extracting software. In fact CAT-tools (e.g. *memoQ*) have this feature with various settings: maximum length words, minimum frequency, sources, etc.[16]

Term bases usually do not contain explanatory remarks or grammatical categories with examples, only terms in strict alphabetical order (even to the detriment of important information, such as US / UK spelling. Similarly, they do not include the long infinitive *to* particle, they hardly ever have definite or indefinite articles preceding nouns, and the creators of a TB are not happy to enter certain types of expressions, such as the Romanian reflexive or phrases (***a se p**lânge – to complain, **a-şi e**xecuta pedeapsa - do penance*, ***a fi b**lestemat – be cursed, **a-şi d**a sufletul - give up one's ghost*). In these cases the 'best option scenario' will indicate that the reflexive construction should be disregarded for the sake of searchability, and the expression should be placed alphabetically under the first word's initial (***e**xecuta pedeapsa*). Thus, for the time being, *hallo.ro* cannot be considered a dictionary as it does not offer a strict alphabetical order: the proof is that there is no difference between *soţie* and *şotie* when one of these terms is searched for.[17] This indicates that the online source is based on the English alphabetical order, which cannot differentiate language specific diacritical marks, only strings of characters. However, this will surely improve in the near future, as it is possible. For instance, once Romanian is selected as the editing language in *Microsoft Office Excel*, we will obtain a correct alphabetical order in Romanian, but if the English spell-checker is activated, we will have a similar alphabetical order to *hallo.ro*.

At this stage we can state again that the combination of printed dictionaries with online dictionaries and TBs opens up multiple possibilities, offering enhanced productivity and quality assurance. It is not acceptable any more to say that online terms are not trustworthy, as large explanatory dictionaries are few clicks away, thus 'cross examination' of a notion, concept or term is very simple. Search engines offer valuable statistics regarding the number of occurrences, and – for instance – *Google Fight*[18]can compare keywords (if we have doubts regarding the spelling or popularity of synonyms). As a result, the best option (term) is secured by a thorough inquiry in various databases. Malicious remarks are often based on the absence of a particular term in a dictionary or online source, or the presence of marginal entries. As the number of words and terms is unlimited, no dictionary or online source will ever contain all the (relevant) entries, and we are sure that many legal terms are still missing from the EU's term base as well. Linguists and experts (*lawyers, attorneys, solicitors, barristers, counsellors, pleasers, proctors, jurists,* etc.) may dedicate their lives to create the ultimate TB or dictionary, which is a never-ending story of term

[16] https://kb.kilgray.com/article/AA-00395/0/Setting-up-and-performing-a-term-extraction-in-memo Q. html, 15. 08. 2015.
[17] http://hallo.ro/search.do?d=en&l=ro&type=both&query=so%C5%A3ie, 15. 08. 2015.
[18] http://www.googlefight.com/, 17. 09. 2015.

BDD-A22791 © 2015 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.216.51 (2025-12-04 20:24:10 UTC)

hunting to exclude irrelevant entries and add new(er) and inventive ones. Translation techniques, such as pure or naturalized borrowings, calques (Molina and Hurtado Albir 2002, 510), Klaudy's (2003, 272–281) antonymous translations (e.g. *la vecinătate – not far from* in Lozinschi 2008, 315) will always provide a fertile soil for them.

Acknowledgement

The research presented in this paper was supported by the European Social Fund under the responsibility of the Managing Authority for the Sectoral Operational Programme for Human Resources Development (*Sistem integrat de îmbunătăţire a calităţii cercetării doctorale şi postdoctorale din România şi de promovare a rolului ştiinţei în societate*), as part of the grant POSDRU/159/1.5/S/133652.

References

Barber, Benjamin R. 1992 (March). "Jihad vs. McWorld". *The Atlantic Monthly* (3): 53–63.

Bennett, S., and Gerber, L. 2003. "Inside commercial machine translation". In *Computers and Translation: A Translator's Guide*, ed. by H. L. Somers. 175–190. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Coteanu, I., Seche, L., and M. Seche. 1996. *Dicţionarul explicativ al limbii române* (2nd ed.). Bucureşti: Univers Enciclopedic.

Imre, Attila. 2013. *Traps of Translation*. Braşov: Editura Universităţii "Transilvania."

Imre, Attila. 2014a. "Jogi szakszövegek és terminológiai adatbázisok." *Hungarológiai Közlemények* (4): 13–23.

Imre, Attila. 2014b. "Ways to enhance legal dictionaries". *Communication, Context, Interdisciplinarity* (3): 519–527.

Imre, A. and Barabás, B. 2015. "Legalese term base". *Studia Universitatis "Petru Maior," Philologia* (18): 68–73.

Kis, B., and A. Mohácsi-Gorove. 2008. *A fordító számítógépe*. Bicske: Szak Kiadó.

Klaudy, Kinga. 2003. *Languages in Translation*. Budapest: Scholastica.

Lozinschi, Smaranda. 2008. *Dicţionar juridic Român - Englez*. Bucharest: Editura Smaranda.

McKay, Corinne. 2006. *How to Succeed As a Freelance Translator*. Lulu.com.

Molina, L., and A. Hurtado Albir 2002. "Translation Techniques Revisited: A Dynamic and Functionalist Approach". *Meta: Translators' Journal* 47(4): 498–512.

Samuelsson-Brown, Geoffrey. 2010. *A Practical Guide for Translators*. Multilingual Matters.

BDD-A22791 © 2015 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.216.51 (2025-12-04 20:24:10 UTC)

Snell-Hornby, Mary. 2006. *The Turns of Translation Studies: New Paradigms or Shifting Viewpoints?* Amsterdam/Philadelphia: John Benjamins Publishing.

Trumble, W. R., and A. Stevenson (eds.). 2002. *The Shorter Oxford English Dictionary* (5th ed., Vol. II). Oxford: Oxford University Press.

Tveit, Jan-Emil. 2009. "Dubbing versus Subtitling: Old Battleground Revisited". In *Audiovisual Translation. Language Transfer on Screen*, ed. by J. Díaz Cintas and G. Anderman. 85–96. Palgrave Macmillan.

BDD-A22791 © 2015 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.216.51 (2025-12-04 20:24:10 UTC)