

Dictionary definitions as text corpora – a phonolexicographic perspective

Włodzimierz Sobkowiak

School of English, Adam Mickiewicz University, Poznań, Poland

0. Abstract

Electronic monolingual EFL dictionary definitions can be used as a textual corpus for use both in lexicography and language pedagogy. A phonolexicographic analysis of a corpus of phonetically transcribed and annotated MEDAL (*Macmillan English dictionary for Advanced Learners*) definitions is presented. Some conclusions are drawn for lexicography and EFL teaching. From the former perspective, definitions in learner's dictionaries can be phonetically more 'user-friendly', i.e. more rigorously controlled in terms of phonetic difficulty of the defining vocabularies implemented (as measured with an L1-sensitive phonetic difficulty Index, PDI) and of the inter-word junctures (sandhi) as well as other suprasegmental phenomena. From the perspective of EFL teaching and learning, definitions can be concorded *qua* corpus to query for phonetically sensitive strings. Such KWIC concordances can then be used in data-driven inductive learning of foreign pronunciation, as well as in exercise and test preparation. The latter can be conducted semi-automatically within the context of rich didactic functionalities of electronic dictionaries. Examples of both lexicographic and didactic benefits of using dictionary definitions as corpora are presented.

1. Background & assumptions

- "In learning the foreign vocabulary for native words, the pronounceableness of the foreign word has a strong determining effect (0.37) depending on the degree to which it conforms to the phonotactic patterns of the native language, even if the word does not have to be spoken" [as quoted by Ellis & Beaton 1993 from Rodgers 1969; my emphasis – WS].
- "The MLD [monolingual learners' dictionary – WS; ...] should be seen first of all as a language learning resource, its compilers as much in the business of language teaching as of lexicography" (Rundell 1988:134).
- Even in the most sophisticated electronic dictionaries for FL learners definitions (as well as example sentences) remain almost completely pedagogically unexploited as a corpus of FL text. And yet, the option of exploiting the dictionary text as a corpus was on the wish-list of on-line OED users in the study of Gray as early as 1986.
- While definition corpora are admittedly inauthentic, in terms of the so-called 'input-enhancement' (see Kettemann 1997) they are actually better than other types because they are necessarily much better controlled on all levels: typographic, linguistic, pragmatic, cultural, etc.
- There is no reason to assume that the acquisition of vocabulary, whether deliberate or incidental, mediated through dictionary definitions is in principle different from one effected in the course of 'ordinary' reading (see Whitcut 1986:119).
- It stands to reason that phonetically difficult definitions will tend to impede the reading and understanding process, the essential part of dictionary lookup, particularly in those learners who continue to vocalise or articulate subvocally in silent reading (hypothesis to be tested).

2. Data – MEDAL definitions

The MEDAL definition files which I obtained from Bloomsbury counted 93042 records altogether, 88495 of which were used in the British edition of MEDAL. Only the latter were used for all calculations. The text of all definitions was first tokenized, yielding 1,054,323 words altogether (15,265 types), and then automatically transcribed phonetically by looking each word up in an electronic word-list. Phonetic Difficulty Index (PDI) tags were also transferred. These originated from the application of a PDI algorithm to the word-list. The PDI algorithm (see Sobkowiak 2004) assigns to each word PD tags coming from a list of 57 (grapho-)phonetic problems characteristic for Polish learners at an intermediate level, such as those illustrated in Table 2 below.

Phonetically transcribed (quasi-SAMPA style) and PDI annotated, MEDAL definitions look like in Table 1:

Table 1. An example of PDI-tagged MEDAL definition (*taster*)

Definition	<i>a small amount of something that is offered so that you can experience it and decide whether you like it or not</i>
Transcription	@ smOl @'m2nt 0v 'sVmTIN D&t Iz '0f@d s5 D&t ju k&n Ik'sp6r6ns It &nd dl'sId 'weD@ ju lIk It O n0t
PDI codes	J 1 gJ N1 EHML N JN1 * L g * C * N NO AJL1 g * * A1 1
Mean PDI	1.3
Number of words	22
Global PDI	28

3. Phonolapsological analysis of definitions – an example

Table 2. MEDAL definition of *discrimination*, mean PDI=3.2, analyzed phonolapsologically

word	PDI value	PDI code	phonetic difficulties and likely mispronunciations
<i>unfair</i>	6	ABEQ13	linking /r/; /eə/; /ʌ/ quality and overnasalized; accent confusion; secondary stress: /ã'fer/
<i>treatment</i>	2	JQ	schwa quality and overnasalization: /tritmēt/
<i>of</i>	2	N1	devoicing; spelling pronunciation; accent confusion: /af/
<i>someone</i>	2	dE	/ʌ/ quality: /'samwan/
<i>because</i>	3	fN1	-au- (spelling pronunciation); final devoicing; accent confusion: /br'kous/
<i>of</i>	2	N1	devoicing, spelling pronunciation; accent confusion: /af/
<i>their</i>	5	cABL1	-ei- (spelling pronunciation); linking /r/; /eə/; /ð/; accent confusion: /zei/, /zer/, /ðeir/
<i>religion</i>	2	JU	schwa quality; /dʒ/ quality: /re'lidʒɪn/
<i>race</i>	0	*	-
<i>or</i>	2	A1	linking /r/; accent confusion: /ɔ/
<i>other</i>	5	AEJL1	linking /r/; /ʌ/ and schwa quality; /ð/; accent confusion: /aze/, /aðe/
<i>personal</i>	5	JKQX1	short and long schwa quality, vowel overnasalization, syllabic /l/, accent confusion: /'personal/
<i>features</i>	5	bJNU1	-ur-; schwa quality; final devoicing; /tʃ/ quality; accent confusion: /fitʃes/, /fitʃers/

4. Main findings

- The most powerful factor responsible for the ultimate phonetic profile of the definitions is the composition and usage of the defining vocabulary adopted for the given dictionary. Frequently used DV words will contribute most to the phonostatics of the entire definition database.
- The average word-weighted phonetic difficulty of definitions is significantly lower than that counted for the English lexicon: 1.52 versus 2.45, respectively. There is little evidence, however, that this happens as an effect of deliberate phonolexicographic control.
- Phonetic difficulty of definitions is partly affected by the prosodic and juncture phenomena: phrase and sentence stress, rhythm, sandhi assimilations, reductions and deletion.
- While it does not appear to be feasible to completely redesign the process of DV selection and implementation (definition writing) to include only phonetically easy words and word clusters in definitions written for learners, better understanding of the phonetic aspects of EFL dictionary definitions among dictionary makers might significantly improve the phonolapsological profile of definitions making them phonetically easier (more 'user-friendly') to learners.

5. EFL application: concordancing definitions for phonetically sensitive strings

Hewings 1993 argues that word-lists and concordanced texts can be an excellent resource in providing teachers and learners with contextualized phonetic practice, by drawing attention to grapho-phonemically difficult rules and items. Some examples follow:

/t+j/ coalescence; PDI<0.6:

bedroom: *a room that you sleep in*

cone: *a cone shape that you put ice cream in and eat*

green: *not yet ready to be eaten*

payphone: *a telephone in a public place that you pay to use*

Linking /r/; PDI<0.6

chasm/crevasse: *a very deep crack in rock or ice*

exactly: *in every way or every detail*

intense: *very great or extreme*

severe/ly: *very strict or extreme*

to have one foot in the grave: *to be very old or ill and likely to die soon*

Schwa-less definitions; PDI<.06

creep by: *if time creeps by it passes very slowly*

lean (adj): *lean meat has very little fat in it*

lean (n): *meat that has very little fat in it*

not a moment too soon: *so late that it is almost too late*

tied up: *if traffic is tied up it is not moving very quickly*

Schwa-heavy definition

client-server: *used for referring to a network (=group of computers) in which each computer is either a client or a server. Clients are the individual computers that run programs or the equipment connected to them such as printers, and servers are the powerful computers that supply the information that makes them work (30 schwas in 51 words, PDI=2.0)*

6. To be studied – a corpus perspective

- The extension of these methods and procedures onto definition and example sentence corpora of other EFL dictionaries.
- Investigating the style of EFL dictionary definitions in comparison with different modes and genres of native and non-native text corpora of English.
- Exploring the didactic potential of dictionary definitions and examples as corpora of native text, on the phonetic, morphological, syntactic, semantic and pragmatic levels.
- Elaborating best methods to offer teachers and learners access to such corpora as an integral part of CALL resources.

References

- Ellis, N.C. & A. Beaton. 1993. "Psycholinguistic determinants of foreign language vocabulary learning". *Language Learning* 43. 559-617.
- Gray, J.C. 1986. "Creating the electronic New Oxford Dictionary". *Computers and the Humanities* 20. 45-9.
- Hewings, M. 1993. "Using a concordancer in writing pronunciation materials". *Speak Out*, January. 11-16.
- Kettemann, B. 1997. "Concordancing as input enhancement in ELT". In B. Lewandowska-Tomaszczyk & P.J. Melia (eds). 1997. *Proceedings of the 1st International Conference on Practical Applications of Language Corpora (PALC)*. Łódź: Łódź University Press. 63-73.
- Rodgers, T. 1969. "On measuring vocabulary difficulty: an analysis of item variables in learning Russian-English vocabulary pairs". *International Review of Applied Linguistics* 7. 327-343.
- Rundell, M. 1988. "Changing the rules: why the monolingual learner's dictionary should move away from the native speaker tradition". In M. Snell-Hornby (ed.). 1988. *ZüriLEX '86 Proceedings* [2nd EURALEX Congress, Zürich 1986]. Tübingen: A. Francke. 127-137.
- Rundell, M. (ed.). 2002. *Macmillan English Dictionary for Advanced Learners (MEDAL)*. Oxford: Macmillan Education.
- Sobkowiak, W. 2004. "Phonetic Difficulty Index". In W. Sobkowiak & E. Waniek-Klimczak (eds). 2004. *Dydaktyka fonetyki języka obcego*. Zeszyt Naukowy Instytutu Neofilologii Państwowej Wyższej Szkoły Zawodowej w Koninie nr 3. Konin: Wydawnictwo PWSZ w Koninie. 102-107.
- Sobkowiak, W. 2006. *Phonetics of EFL dictionary definitions*. Poznań: Wydawnictwo Poznańskie.
- Sobkowiak, W. et al. 2006. "What are dictionary definitions good for?". In K. Dziubalska-Kończyk (ed.). 2006. *IFAtuation. A life in IFA. A festschrift for professor Jacek Fisiak on the occasion of his 70th birthday*. Poznań: Wydawnictwo Naukowe UAM. 655-684.
- Whitcut, J. 1986. "The training of dictionary users". In R.F. Ilson (ed.). 1986. *Lexicography: an emerging international profession* [Fulbright Colloquium, London 1984]. Manchester: Manchester University Press. 111-122.

Notes: