Phonetics of EFL dictionary definitions
(an excerpt)

Włodzimierz Sobkowiak

In the following I will simply assume that the properly annotated definitions function as a corpus-like resource within the dictionary. This means that they can not only show phonetic information (PDI rating, for example) to the user, if so toggled, but also that they can be interactively queried for text meeting some user-defined phonetic criteria. This is of course not unlike the common 'definition search' in e-dictionaries, which – in some advanced products – provides the user with a variety of Boolean-combinable search options: part-of-speech, grammar, style, frequency, illustration, etc. No phonetic criteria have so far been implemented in 'definition searching' of any dictionary known to me, however[1].

Below, a number of the many thinkable (and didactically useful) phonetic search results through MEDAL definitions are exemplified. Typographically speaking, with one exception, they are not KWIC-style aligned, but this is of course easy to achieve, should one want to display search results in this way for user's convenience.

Should the user of the phonolapsologically enabled dictionary wish so, there is no reason why a KWIC concordance output should not display hits in phonetic transcription, the trick used by Neuhaus 1995 in his *Phonetic concordance to Daniel Jones*. The result might look somewhat like the following (definitions with /-ɪ'eɪʃn/ and PDI<1.5; see also section 7.5. below), a *phoncordance* of sorts:

- juzd In strit n4mz D@ 'rItn @,briv**I'4Sn** Iz &v
- @ 'medIkl 'tritm@nt In wIC ,r4d**I'4Sn** Iz juzd fO 'tritIN &n
- t@ juz ,r4d**I'4Sn** &z @ 'medIkl 'tritm@nt 1z
- m4k smOl kwIk 'muvm@nts rem Iz &n @,briv**I'4Sn** fO 'r&pId 1 'muvm@nt
- If &n @,briv**I'4Sn** OR @ 'sImbl st&ndz fO
- t@ trit fud wID ,r4d**I'4Sn** @ t1p 0v 'en@JI In 'Od@ t@

The full potential of phonetic annotation of definition text (as well as any other type of EFL text) can only be released when phonolapsological annotation of the PDI type is used. This section will be further divided into: (a) searches based on PDI values, (b) searches based on transcription and PDI monograms, (c) searches based on PDI polygrams, and (d) searches based on the "words commonly mispronounced" list in Sobkowiak 1996a.

7.2. Searches based on transcription and PDI monograms

The variety of thinkable queries involving transcriptional strings and the 57 PDI monograms (unary codegrams, 1-grams) is practically unlimited. Notice that the query terms can be used

---

[1] This is of course not to say that no phonetic searches whatsoever are possible in e-dictionaries. MEDAL itself has a rather extensive *SoundSearch* facility, which makes it possible to use phonetic (mostly segmental) information to select keywords. LDOCE has recently added this functionality to the CD-ROM version, too.

singly or in combinations of arbitrary number (combinations of PDI codes within words, however, are treated separately as polygrams). In the following I can only offer a tiny selection focused on the needs of the Polish EFL learners, as they are recognized in the literature and my own teaching experience. The design of the PDI measure is, after all, explicitly guided by this context. No attempt will be made to be exhaustive in any sense, say in presenting all definitions meeting some predefined criteria. Similarly, I will not try to offer a full exposition of each phonetic problem as we go along, as there are other excellent sources devoted entirely to this one task. On the other hand, I believe a short introduction into the nature of the phonetic difficulty under query is in order.

7.2.1. Definitions with sandhi /t+j/ coalescence; PDI<0.6

While no sandhi phenomena are tagged in the current implementation of my PDI algorithm (with the exception of linking /r/), it is still possible to search for some of these in a fully transcribed definition. One example is /t+j/ coalescence, a fast-speech phonetic process posing specific pronouncing difficulties to (Polglish) EFL learners. There is no coalescence rule in Polish comparable to the one at hand. Not surprisingly, Poles tend to pronounce the two segments separately (as they are spelled) until they develop a certain measure of phonostylistic automaticity, which is in turn conditioned by fluency. Before this stage, the rule, which seems so thoroughly automatic to native English speakers, quite simply does not apply in Polglish. Some definition examples containing the relevant strings, and yet relatively phonetically easy (PDI<0.6) are listed below:

- bedroom: *a room tha**t y**ou sleep in*
- cone: *a cone shape tha**t y**ou put ice cream in and eat*
- green: *no**t y**et ready to be eaten*
- payphone: *a telephone in a public place tha**t y**ou pay to use*

7.2.2. Definitions with many cases of aspiration

Sandhi phenomena are not the only ones which are untagged by the PDI algorithm, and yet searchable through phonetic transcription. The algorithm is almost completely phonemically oriented, so specific types of allophonic variation are not captured. In this context aspiration comes most naturally to mind as the allophonic process in terms of difficulty and amount of didactic care in the area of EFL. As Polish exhibits no aspiration of stops, Polglish learners at all levels of pronouncing proficiency find it difficult to apply it consistently and according to the several conditioning factors which come into play: stress, segmental context, speech tempo, etc.). The following examples of MEDAL definitions were selected because in each one there are a few tokens of aspirated plosives in the canonical context, i.e. immediately before a strongly stressed vowel.

- rubbing: *a **p**icture that you make by **p**utting **p**aper over a **p**attern on stone or metal and rubbing it with something such as a **p**encil*
- tiger: *a large Asian wild animal that has yellowish fur with black lines and is a member of the cat family The female **t**iger can be called a **t**iger or **t**igress A young **t**iger is called a **t**iger cub*
- DDT: *a poisonous **ch**emical used for **k**illing insects that destroy crops It also **k**ills animals and is dangerous to humans so is no longer allowed in many **c**ountries This **k**ind of **ch**emical is **c**alled a pesticide or an insecticide*

7.2.3. Definitions with pre-sandhi /t/ deletion; PDI<0.6

Another phonostylistically motivated process of quite some importance in the context of EFL pronunciation is word-final alveolar stop deletion. While it occurs in a variety of contexts, only pre-sandhi elision out of word-final /-st#/ and /-nt#/ clusters[2] is illustrated here, with the following word beginning in a consonant. The teacher could easily use this phonetic information, normally buried under thick layers of spelling, to make search-based dictionary exercises enforcing on the learner the oral rendition of definition strings.

Table 1. Pre-sandhi /t/ deletion; PDI<0.6

keyword	definition (aligned on /-Ct_C-/)
?	*waste disposal**
mincemeat	*minced meat*
?	*iced tea**
straight (adv)	*in an honest way*
dilly-dally	*to waste time*
gallop	*an extremely fast speed*
immediately	*most directly*
at the last moment/minute	*when it is almost too late*
go out	*to be sent by post*
vitally	*in an extremely important way*
Doric	*built in a plain ancient Greek building style*

* This definition appears in the original Bloomsbury file, but is unattested on the MEDAL CD-ROM

7.2.4. Definitions with cross-sandhi stop geminates

One of the contexts where stop gemination is possible, or indeed required, in English is the word-boundary one, when stops appear on both sides of it. (Mis)guided by spelling, as well as by the strength of their native tongue phonotactic habits, learners often tend to release the first stop, i.e. the word-final one, rather than correctly produce one prolonged segment. It is easy to scan MEDAL definitions for such sequences and extract relevant strings. Most of the cases found are of course those of voiceless plosives: /-tt-/ 4864, /-kk-/ 503, /-pp-/ 267. Of the voiced stop geminates, only /-dd-/ appears with appreciable frequency of 1271. The following are the PDI-easy examples containing all six strings in definitions of medium length.

- dead to the world: *sleeping in a way that makes it difficul**t t**o wake you*
- cast (verb): *if a sna**ke c**asts its skin it slides out of it*
- put in: *if a shi**p p**uts in it stops at a port*
- drip (verb): *if a liqui**d d**rips it falls in very small drops*
- machine: *a piece of equipment that does a particular jo**b b**y using electricity steam gas etc*
- grunt (verb): *if an animal such as a pi**g g**runts it makes a low sound*

[2] This is the most common preceding segmental context for this process (see Sobkowiak 2001:281)

7.2.5. Definitions with prevocalic definite article; PDI<0.9

Another inherently sandhi phenomenon of English phonotactics is the strengthening of the schwa vowel to a high-front quality in the definite article when it precedes a vowel-initial noun. Until fairly advanced levels of EFL pronouncing proficiency it is simply too hard for learners to produce the two most phonetically difficult sounds of English in one word (/ðə/) and control for phonotactic and stylistic variation at the same time (the 'strength' of the article is also conditioned stylistically, of course). Reading practice material with appropriate prevocalic *the* contexts is therefore in demand, and there is every reason to use phonetically transcribed definitions in this capacity.

While the article was uniformly transcribed as /ðə/ by the lookup algorithm, for reasons explained in section 6.1., it is easy enough to retrieve only those definitions where it occurs prevocalically (and, if need be, adjust the transcription accordingly). The following is a tiny sample of the 3955 definitions meeting this condition. I have selected the PDI-easiest ones (PDI<0.9).

- scouting: ***the*** *activities that boy scouts take part in*
- loitering: ***the*** *offence of standing or waiting in a public place so that it looks as if you might commit a crime*
- distance (noun): ***the*** *amount of time between two points in time*
- club (noun): *a place you go to in* ***the*** *evening to dance and drink*
- initiative: ***the*** *ability to decide in an independent way what to do and when to do it*
- disturb the peace: *to commit* ***the*** *illegal act of behaving in a noisy way in public especially late at night*
- pudding/sweet (noun): *a sweet food that you eat at* ***the*** *end of a meal*
- classically: *in a style based on* ***the*** *Ancient Greek and Roman styles*

7.2.6. Definitions with linking /r/; PDI<0.6

Linking /r/ was marked as a potentially realized segment in the original OALDCE word-list file, which I used for my lexical database. Once the /r/-final words are entered into running transcription, of course, some of these original R's must be dropped in British English. Those which remained in MEDAL definitions can now be used for querying the corpus. As before, only the phonetically easiest definitions appear below:

- chasm/crevasse: *a very deep crack in rock o****r i****ce*
- exactly: *in every way o****r e****very detail*
- intense: *very great o****r e****xtreme*
- severe/ly: *very strict o****r e****xtreme*
- to have one foot in the grave: *to be very old o****r i****ll and likely to die soon*

7.2.7. Definitions without schwa; PDI<0.6

Learners of English pronunciation quickly discover, both from experience and explicit instruction, that schwa is, textually speaking, the most common English phoneme. Most of the time, they are also painfully aware of their problems trying to pronounce the fuzzy, lax central vowel. Nineteen in twenty MEDAL definitions contain this ubiquitous sound, so there is no shortage of example sentences to practice. But maybe the teacher would like to have only those few definitions without schwa? It is easy enough to filter these in, with the additional criterion of PDI<0.6:

- 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*

7.2.8. Definitions with many schwas

Schwa is pervasive in English and in MEDAL definitions not only in the percentage of definitions in which it occurs but also in its incidence within any particular definition. 65248 definitions out of the total of 88495 (73.7%) contain as many as seven schwas in them. Should there be a need to practice the pronunciation of schwa in a schwa-dense context, the corpus of definitions can provide a lot of material. The following are the five definitions in MEDAL with the highest incidence of schwa (the count for each is given in brackets). Notice that most of them are relatively easy, with one (*standing order*) even below the global PDI mean of 1.52. The schwa-count record holder in MEDAL is the definition of *client-server*, which has 30 schwas in 51 words, i.e. on average one schwa in 1.7 words.

- branch: *a part of the government with a particular responsibility The three branches of government are the legislative branch (=the parliament that makes laws), the executive branch (=the president or ministers who govern according to those laws), and the judicial branch (=the judges and courts who make certain that the laws are used correctly)* (27 schwas, PDI=2.3)
- cabinet: *a group of advisers chosen by the leader of a government In the British political system members of the cabinet are called cabinet ministers and each minister is usually the head of one of the main government departments such as Health or Defence* (24 schwas, PDI=1.8)
- 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, PDI=2.0)
- church: *a building that Christians go to in order to worship Traditional churches usually contain an altar and long wooden seats facing the altar called pews The place where the priest or minister stands to talk to the people is called a pulpit A religious ceremony that takes place in a church is called a service* (25 schwas, PDI=1.7)
- standing order: *an instruction that you give a bank to take a particular amount of money out of your account on a particular day usually each month to pay a person or organization for you A direct debit is a similar arrangement except that the amount can change and is decided by the person you are paying* (25 schwas, PDI=1.5)

7.2.9. Definitions with many word-final voiced obstruents

Word-final voiced obstruents are almost as frequent in the English lexicon (CUV3) as schwas. In the pronunciation learning and teaching practice, then, they are bound to come up equally often. Both are phonetically problematic to (Polish) learners: schwa in terms of fixing the proper tongue position (timbre), tenseness and length; word-final voiced obstruents in terms of voicing. Final obstruent devoicing is one of the strongest and most pervasive phonological rules of Polish; no wonder that keeping the right level of voicing in the relevant English contexts is very hard to learners. The dynamism of interlanguage usually leads from total

devoicing first, through exaggerated voicing (with a supporting final schwa in some cases), to the correct level of voicing in the advanced stages of competence development.

Final voiced obstruents occur in almost 37% of CUV3 words, almost 30% of MEDLIST words, and almost 85% of MEDAL definitions. Like in the case with schwa above, there is thus no shortage of examples to practice, but it may be useful to extract those definitions which contain most word-final voiced obstruents. Using the PDI code <N> one can do it for all eight such phonemes in English at once. The result is listed below.

- reserve currency: *a currency (=type o**f** money use**d** in a particular country) that i**s** considere**d** a**s** strong an**d** reliable an**d** i**s** use**d** a lot in international tra**de** National banks keep lar**ge** store**s** o**f** reser**ve** currencie**s** which inclu**de** the US dollar the poun**d** an**d** the Japane**se** yen* (19 cases)
- tide: *the way that the level o**f** the sea regularly rise**s** an**d** fall**s** during the day The ti**de** come**s** in or rise**s** then it turn**s** an**d** goe**s** out or fall**s** an**d** then turn**s** again At its highest point the ti**de** i**s** in an**d** it i**s** high ti**de** an**d** at its lowest point the ti**de** i**s** out an**d** it i**s** low ti**de*** (24 cases)
- two-two: *a university degree in the UK that i**s** at the lower en**d** o**f** the secon**d** level Degree**s** are divide**d** into three level**s** The first level i**s** calle**d** a first the secon**d** i**s** divide**d** into two level**s** a two-one an**d** a two-two an**d** the thir**d** i**s** calle**d** a thir**d** A two-two i**s** also calle**d** a lower secon**d*** (22 cases)

7.2.10. Definitions with many voiced apico-dental fricatives (/ð/)

The trick with filtering through only definitions containing many occurrences of an arbitrary phonetic difficulty for practicing is of course very easy to apply to any PDI codegram. Not to labour this point, I will list below one more example catch: definitions with many /ð/'s, that veritable tour de force of English pronunciation.

- North-South Divide, the: ***th**e difference in attitudes or situation between people in **th**e nor**th**ern and sou**th**ern parts of a country In **th**e US **th**e south is generally more conservative **th**an **th**e north while in **th**e UK **th**e north is generally poorer **th**an **th**e south* (12 cases)
- Mason-Dixon Line, the: ***th**e border between **th**e two states of Maryland and Pennsylvania in **th**e US **th**at traditionally divides **th**e north of **th**e US from **th**e south In **th**e past **th**is was **th**e border between **th**ose sou**th**ern states **th**at supported slavery and **th**e nor**th**ern ones **th**at did not* (16 cases)

7.2.11. Definitions with secondarily stressed words; PDI<0.8

One of the few non-segmental symbols appearing in the original OALDCE word-list are stress signs. Both primary and secondary stress are marked. Lexical stress of English is hard to foreign learners, mostly because they find it hardly at all predictable. Of the two stress levels, it is naturally the secondary which is the harder to acquire, if only because it is auditorily much less salient. It is easy enough to find definitions containing secondarily stressed words in MEDAL:

- coded: *expressed in an ˌindirect way*
- dub: *an ˌelectronic type of music based on reggae music*
- not/never miss a trick: *to notice every ˌopportunity and use it*
- referee (v): *to be a ˌreferee in a game*

- take your leave: *to say ˌgoodbye*

7.2.12. Definitions with selected word-stress patterns

Naturally, it is not only the secondary stress that can be searched for. Practically any of the twenty or so lexical stress patterns in English can be queried through the phonetic transcription field of the definition database, from the simplest disyllabic ones, say iambic, to the most complex. To show some examples of the latter, the following is a selection of definitions with words showing stress patterns extending over five and more syllables.

- autobiographical: *relating to someone's life or ˌautobi'ography*
- radiotherapy: *a medical treatment for cancer using radiation Someone who is trained in ˌradio'therapy is called a ˌradio'therapist*
- biodegradable: *ˌbiode'gradable substances can be broken into very small parts by bacteria so that they are not harmful to the environment*
- radium: *a ˌradio'active element used in the treatment of cancer*
- TB: *ˌtubercu'losis: a serious infectious disease that mainly affects the lungs*
- terrestrial: *existing on the earth or happening on the earth instead of in the sky or sea Things that are from other planets are ˌextrater'restrial*
- hallucinatory: *similar to a ˌhalluci'nation*
- transcontinental: *a ˌtransconti'nental flight railway or road goes from one side of a continent to another for example across Asia or America*
- vet (noun): *a doctor for animals Vet is short for 'veterinary surgeon*

7.2.13. Definitions with long words; PDI<1.5

The issue of long words was raised a few times already in the analytic part of this book. Their frequency in MEDAL was provided, as well as a few observations on its affinity to compounds (code <a>) and secondary stress (code <3>). It was also found that only relatively short secondarily stressed words of DV are selected for definitions. This does not mean, of course, that a teacher or learner desiring to practice the pronunciation of really long words (> 5 syllables) in context would find little of use in MEDAL definitions. Even once *responsibility* and its plural are excluded (by far the most frequent code <2> word), there is enough material in definitions with a relatively low PDI value (<1.5):

- polio: *a serious infectious disease that mostly affects children It destroys muscle and can cause paralysis (=lack of movement in your arms and legs) Polio is a short form of the more technical word* ***poliomyelitis***
- biodegradable: ***biodegradable*** *substances can be broken into very small parts by bacteria so that they are not harmful to the environment*
- biography: *a book that someone writes about someone else's life A book that someone writes about their own life is called an* ***autobiography***
- radio wave: *an* ***electromagnetic*** *wave that radio signals can be sent on*
- intercommunicate: *if two rooms* ***intercommunicate*** *there is a door in the wall that separates them*
- indistinguishable: *if two things are* ***indistinguishable*** *you cannot see any difference between them*
- half-life: *the amount of time that is needed for a substance to lose one half of its* ***radioactivity***

7.2.14. Definitions with *-ic* in disyllabic-plus adjectives; PDI<1

Multisyllabic adjectives ending in *-ic* present a particular difficulty to EFL learners: the suffix is one which forces stress change, e.g. *'alcohol → alco'holic*. Such shifting stress is completely alien to Polish learners with their habit of stressing the penultimate syllable in almost all words. Many Polglish learners learners are also quite suspicious of English paroxytones such as *alcoholic*, as they seem to be deceptively similar to the Polish standard. With the single <5> code we can capture all these words in their natural context in definitions. Of course, considering the relative frequencies of the 115 MEDAL code-<5> words, one can expect definitions with the most common ones to predominate. Such is indeed the case, as seen in the following examples, with PDI<1.

- bookmark (verb): *to mark an Internet website in an **electronic** way so that you can easily find it again*
- you bet: *used for saying yes in an **enthusiastic** way*
- radio wave: *an **electromagnetic** wave that radio signals can be sent on*
- audiotape: ***magnetic** tape on which sound can be recorded*
- ironically: *in an **ironic** way*
- at some point: *at some moment in time that is not made **specific***

7.2.15. Definitions with three English centring diphthongs; length <13 words

English centring diphthongs are hard to learners of various mother tongues, including Polish. The source of this difficulty is both articulatory and acoustic. As far as the former is concerned, it involves: (a) the qualities of both vowels in the diphthong, (b) their linking into one tautosyllabic unit, (c) effects of stress or lack of it, (d) phonostylistic effects (gliding, monophthongization, etc.). It is clear that it may be advantageous to be able to collect all three centring diphthongs in one definition in order to address those potential problems in close proximity, appearing in stylistically natural strings. Under the present PDI tagging scheme it is possible to do this both by reference to the transcription of the specific phonemes and by using the PDI codes directly (<B, C, D>).

- uncharacteristic: *not typical of someone's us**ua**l behav**iou**r or performance and th**ere**fore surprising*
- true: *r**ea**l or act**ua**l especially when comp**are**d with how something seems to be*
- luxuriant: *lux**uria**nt h**air** is thick and healthy*
- idiosyncratic: *id**io**syncratic tendencies are unus**ua**l or strange and not sh**are**d by other people*
- strike pay: *money paid to workers by th**eir** un**io**n d**ur**ing a strike*
- odds and sods: *v**ariou**s people or things that are not important or val**ua**ble*
- wear away: *if a feeling w**ear**s away it grad**ua**lly disapp**ear**s*

7.2.16. Definitions with hard sandhi consonant clusters: /s_ð/ and /z_ð/; PDI<0.7

If it is indeed true that, apart from their inherent segmental difficulty, "the difficulty of /θ/ and /ð/ lies [...] in their combination with other fricatives, especially /s/ and /z/" (Cruttenden 1994:168), it might be useful to extract from among definitions only those which contain just the strings in question. As PDI is not well-suited for getting at sandhi phonetic phenomena,

recourse can be taken to phonetic transcription directly. Among the relatively PDI-easy MEDAL definitions some do contain the troublesome clusters, usually with /ð/-initial function words following the fricative-final content ones:

- dripping: *fat and ju**ice th**at is produced by meat when it is cooked*
- mitigating circumstances: *fact**s th**at help to explain a crime or mistake and make it seem less bad*
- payphone: *a telephone in a public pla**ce th**at you pay to use*
- cram: *if people cram a pla**ce th**ey fill it completely*
- unvarnished: *expressed in a very direct way that giv**es th**e true facts*

7.2.17. Definitions with hard sandhi consonant clusters: /dʒd_ð/ and /tʃt_ð/; PDI<1.5

The voiced apico-dental fricative beginning the definite article in English is involved, by the sheer textual frequency of the word, in many other heavy sandhi clusters of particular difficulty to EFL learners. One of them arises when a /ð/-initial word (usually the definite article) follows a code <F> word, i.e. past-tense or past-participle form of a verb ending in one of the two English palato-alveolar affricates: /tʃ/ and /dʒ/. As these are, phonetically speaking, bisegmental, a four-item cluster originates requiring lightning-fast and very accurate adjustment of tongue shape and position for correct production: from alveolar (/t/), through palato-alveolar (/ʃ/), back to alveolar (/t/), and finally to apico-dental (/ð/). On a completely different, morphophonemic level, (pre-)intermediate learners often have doubts whether they should interpret the suffixal allomorph as identical to that where it is attached to /t, d/-final verbs, like in *connected* and *intended*, hence */'dæmɪdʒɪd/ or */'ri:tʃɪd/, for example. All in all, the cluster is indeed among the hardest sandhi strings to produce for learners, and it would be nice to be able to use the definition corpus to extract relevant material for practice. The word-final cluster can be identified both via transcription and via the <F> code, the /ð/-initial word – via transcription (incidentally, very few words in English begin orthographically with *th-* which is <u>not</u> phonetically realized as an apico-dental fricative: *Thames, Thai, thyme*). Low PDI (<1.5) definitions are exemplified below.

- write-off: *a vehicle that is so badly dama**ged th**at it cannot be repaired*
- down: *if you are down for a particular job it has been arran**ged th**at you will do it*
- engaged: *if two people are enga**ged th**ey have formally agreed to get married*
- witness (verb): *to watch someone sign an official document and then sign it yourself to state that you have wa**tched th**em*
- minor (noun): *someone who has not rea**ched th**e age at which they are legally an adult*

7.2.18. Definitions with strings prone to vowel overnasalization; PDI<0.7

One of the recalcitrant phonetic problems of English to Polish learners is the articulation of pre-nasal vowels. In the following tautosyllabic consonantal context (thus: -VnC), but especially when the consonant is a fricative, the vowel coalesces with the nasal in Polglish to form a nasal(ized) vowel, quite similar to the timbre-compatible Polish one. A few examples of this process were presented above, where specific definitions were phonolapsologically analyzed in-depth: /'gavermẽ(n)t/, /ã'fer/, /'tritmẽ(n)t/. Using the PDI <Q> code it is possible to retrieve such cases without reference to specific phonemic strings. Most examples below

show overnasalized schwas because this vowel is the most frequent generally and in the pre-nasal context. A few definitions with PDI<0.8:

- stand out in a crowd: *to be very differ**en**t and easy to notice*
- folksy: *informal in a fri**en**dly way that makes you feel relaxed*
- intervene: *to delay an ev**en**t or make it difficult to do*
- meet: *to play ag**ain**st an oppon**en**t in a game*
- plastic: *a plastic subst**an**ce can be b**en**t into any shape and will keep that shape*
- gallant: *a man who is gall**an**t treats women in an extremely polite and helpful way*

7.2.19. Definitions with /n/ prone to palatalization

One of the Polglish phonetic difficulties identified by Szpyra and others (Szpyra-Kozłowska, Frankiewicz & Gonet 2002) but not explicitly coded by my PDI algorithm is the tendency to palatalize the English alveolar nasal to a palatal one (/ɲ/) in pre-palatal contexts, i.e. mostly the high-front vowel, the palatal glide /j/ and the /ɪə/ diphthong, in which the /ɪ/ vowel is customarily raised and fronted to the Polish /i/ quality. Even though this problem is not directly captured with a PDI code, it is possible to scan MEDAL definitions for relevant transcription strings. In the following examples there are at least two such contexts in each definition.

- stringer: *a journalist who reports local **new**s for one or several **new**spapers, magazines or **new**s organizations*
- critically: *by people whose job is to give their opi**ni**on of things such as **new** books, plays or films*
- ingenious: *an inge**ni**ous plan, piece of equipment, etc uses **new** and clever ideas*
- offside: *the side of a vehicle that is **nea**rer to the middle of the road The side that is **nea**rer the edge of the road is the **nea**rside*
- on bended knee: *with one **knee** or both **knees** on the ground*
- ingenious: *someone who is inge**ni**ous is good at inventing things and solving problems in **new** ways*
- table d'hôte: *a me**nu** in a restaurant that offers a fixed price for meals A me**nu** on which each dish has a separate price is an a la carte me**nu***
- conclave: *a secret meeting of cardinals se**ni**or priests in the Catholic church in which they elect a **new** Pope*
- Euro-: *relating to the European U**ni**on sometimes added to nouns to make **new** words*
- nuclear-free: *a **nu**clear-free country or area does not allow **nu**clear power or **nu**clear weapons to be produced or used within its borders*
- ideogram: *a written sign or symbol such as those used in Chi**ne**se and Japa**ne**se writing that represents an idea or a thing*

7.2.20. Definitions with dark /l/

The other of the contextually conditioned EFL pronunciation problems identified by Szpyra et consortes, but not captured directly by my algorithm is the 'dark /l/, constrained phonotactically to occur before consonants only. This allophone of /l/ does not exist in current standard Polish, so naturally learners tend to transfer the 'clear' /l/ quality into their English. It is easy enough to retrieve from MEDAL definitions with several tokens of 'dar'k /l/, as the sound is quite frequent in running English text. The two definitions below have the maximum number of 'dark' /l/ tokens.

- prefect: *in some schoo**l**s in the UK an o**l**der student who contro**l**s the activities of younger students and he**l**ps them to obey the ru**l**es*
- ethnocentric: *showing a failure to recognize that other peop**l**e's cu**l**tures are a**l**so important and valuab**l**e*

7.2.21. Definitions with syllabic sonorants; PDI<0.8

Using phonetic transcription to retrieve words containing syllabic sonorants is counter-productive because: (a) a disjunction must be formulated (/n/ OR /l/), (b) control over the locus of the segment is difficult (word-final or not). With the PDI codes, both problems disappear: all syllabic sonorants can be obtained in either of the two contexts (codes <X> and <Y>). The value of PDI can be controlled independently, of course, as in the examples below, all very easy phonetically.

- demon: *an ev**il** spirit*
- easy on the eye / ear: *nice to look at / list**en** to*
- to eat like a bird: *to eat very litt**le***
- fly: *to trav**el** by plane*
- narcotic (adj): *ab**le** to make you feel less pain and help you sleep*
- vitally: *in an extremely import**an**t way*

7.2.22. Definitions with British/American phonetic differences; PDI<0.6

The linking /r/ phenomenon, PDI coded <A>, and treated in a previous section, is of course one of the classical symptoms of British English. Logically, then, all words coded <A> are also coded <1>. This code pair is one of the most highly correlated in CUV3 and MEDLIST alike. The implication, however, works in one direction only: from <A> to <1>; obviously, not all words pronounced differently in Great Britain and America need contain the final /r/. Thus, the <1> code is a perfect example of an aggregate PDI monogram: with one symbol it captures a number of phonetic differences between the two main dialects of English, differences which may be thoroughly confusing to learners, hence – lapsogenic.

Because the <1> code is so prevalent in MEDAL definitions, it is easy to get enough practice material from them with rather stringent conditions as to the PDI value and length. The definitions below are all PDI<0.6 and length>6 words. The Am-Br differences captured by the algorithm are bolded.

- weevil: *an insect that eats pl**a**nts and can destroy cr**o**ps*
- chasm/crevasse: *a very deep crack in r**o**ck or ice*
- fly (verb): *if time flies it seems to p**a**ss very quickly*
- soundly: *if you sleep soundly you sleep well and it is n**o**t easy to wake you*
- drip (verb): *if you drip a liquid you make it f**a**ll in very sm**a**ll dr**o**ps*
- sit in: *if people sit in they take pa**r**t in a sit-in*
- jotter: *a sm**a**ll book in which you write notes*

7.2.23. Definitions with vocalic hiatus; PDI<1.3

Word-internal vocalic hiatus in Polish is resolved in one of two ways: (a) by inserting a glide of similar timbre to the first vowel, i.e. /w/ after the high back vowel or /j/ after the high front vowel, or (b) by inserting a glottal stop, regardless of the vowel quality. Polish learners, even

those of fairly high pronouncing proficiency, tend to transfer these habits to English, rendering *situation* as /-uweɪ-/ or /-uʔeɪ-/, respectively. This word is the most frequent of the <S>-coded ones in MEDAL, with 2311 occurrences. Other frequent words include: *association, abbreviation, radiation, punctuation, pronunciation, aviation,* all coded also with <T> for post-alveolar fricatives, the <ST> bigram being one of the most highly internally correlated. However, hiatus can also occur alone, as in: *permeates, menstruates, liaise, mediator, graduated, evacuating, gladiator, radiator, radiate*. The following definitions with <S>, which <u>do not</u> contain *situation*, all meet the condition of PDI<1.3:

- valedictorian: *the student in a high school who has the best marks and usually makes a speech at a grad**ua**tion ceremony*
- period: *the time about once a month when a woman who is not pregnant menstr**ua**tes (=loses blood)*
- radiate: *if lines paths or roads rad**ia**te from a central point they spread out from it*
- permeate: *if an attitude or feeling perm**ea**tes something you can feel or see its influence clearly in every part of that thing*

7.2.24. Definitions with strings prone to word-internal regressive voicing; PDI<2

The Polish tendency to regressively assimilate voicing in sandhi contexts was discussed in section 6.9.3. The rule works even more pervasively within words. No obstruent cluster can be voice-heterogenous in Polish, hence voiceless-voiced clusters in English tend to be pronounced voiced throughout. One common context where this happens is coded PDI <O>: the *mis-* and *dis-* prefixes. The latter is especially frequent in MEDAL, with such word types as: *dishonest, dislike, disagree, disappointed, disappear, disabled, disadvantage*. Definitions suitable for practice of this aspect of English phonotactics include:

- disgust (verb): *if something very unpleasant **dis**gusts you it makes you feel physically ill when you see smell or taste it*
- disband: *if a group of people **dis**bands or is **dis**banded its members stop working together*
- disdain: *if you **dis**dain to do something you refuse to do it because you think it is unimportant*
- misbehabe: *if a child **mis**behaves or they **mis**behave themselves they behave badly and annoy or upset people*
- misdirect: *if a judge **mis**directs a jury he gives them the wrong information*
- misguided: *a **mis**guided idea or action is based on judgments or opinions that are wrong*

7.2.25. Definitions with strings prone to word-internal regressive devoicing

The other side of Polish obstruent cluster voice uniformity rule is, as might be expected, regressive devoicing in Polglish in relevant contexts. The <R> code was devised to partly capture such cases. The word-internal voiced obstruent + /s/ appears to be especially problematic to Polish learners. Using PDI coding, it is possible to extract both those words and definitions containing this code only (<R> monograms) and those where the code is only one in a higher order codegram (polygram). The examples of the former in MEDAL include: *website, abstract, websites, midst, pigsty, abscess,* of the latter: *medicine* <RX>, *baseball* <aR1>, *windscreen* <aR>, *landscape* <aR>, *obsession* <JRTX>, *lobster* <AJR1>. The most phonetically difficult <R>-coded polygrams in MEDAL appear to be: *loudspeaker* (PDI=6), *bobsleigh* (PDI=5), *misbehave* (PDI=5), *misbegotten* (PDI=5) and *obsessions* (PDI=5). The definition corpus can then be queried with these orthographic strings directly.

7.2.26. Definitions with intervocalic /h/

As noted a few times above, one of the common Polglish mispronunciations of the English glottal fricative is its velarization, on the Polish model, to /x/. Another problem, in intervocalic contexts, is the voicing of the incorrect velar to /ɣ/, this time mainly in response to the universal tendency to economize articulatory effort (stopping vocal fold vibration for /h/, only to resume it immediately afterwards, is costly for the speech processor). By reference to phonetic transcription of MEDAL definitions (and the PDI code <V>) one can extract exactly those words and definitions which contain intervocalic /h/. The most frequent lexical items with such a phonemic string in MEDAL include (in alphabetical order): *ahead, alcohol, behave, behind, perhaps*, with their inflections and derivatives. Definitions of PDI<1.1 for practice:

- furtive: *be**h**aving in a way that makes people think you do not want to be noticed*
- vapour trail: *the long white line seen in the sky be**h**ind a plane*
- back onto: *if a building garden etc backs onto a place that place is directly be**h**ind it*
- hole-and-corner: *done in a secret and per**h**aps dishonest way*
- lap (verb): *to pass someone else who is competing in a race when you are a**h**ead of them by a whole lap*

7.4. Searches based on the "words commonly mispronounced" list in Sobkowiak 1996

We have thus reached the limit of the search method whereby phonetic transcription and PDI codes, singly or in combination, are used to scan dictionary definitions for tokens of needed phonetic difficulties. This is, naturally, not the ultimate end of phonetic search possibilities, however. Once phonetically difficult words or strings are identified from other sources, they can be used for searching the corpus in their standard orthographic form. This is of course what is now a standard feature of learners' electronic EFL dictionaries, the 'whole text' search. Some e-dictionaries even provide a search menu where the user can (among other things) constrain the search to certain elements of the entry's microstructure, such as the definition text, for example. MEDAL is one of the dictionaries which has a well-developed menu of this sort. With the whole mechanism of PDI in place, however, the searches are all the more focused and flexible. In the following I present some search examples based on the most common "words commonly mispronounced".

We could, of course, take the first couple of words from Appendix 5, i.e. those with the highest MEDAL frequency, starting with the ubiquitous *people*, and look them up in the definition corpus. This would be more or less what an ordinary user of MEDAL on CD-ROM could also do, taking the search input from my 1996 book. The definition database, however, allows us to run searches of higher sophistication. I decided to scan it for tokens of words which are both fairly difficult (PDI>3) and quite frequent in MEDAL (>100 occurrences). Additionally, that definition was selected in each case which contained the highest incidence of the given lexical item, with the lowest PDI value as a subsidiary criterion. The definitions are listed in the frequency order of the words, which are bolded.

- touch football: *an informal type of American football in which players do not tackle each **other** (=knock each **other** to the ground) but they can touch each **other***
- bustle (noun): *something that women wore round **their** waists in the past to hold **their** skirts away from **their** bodies at the back*

- transducer: *a piece of equipment that gets **power** from one source and then changes that **power** so it can be used by another system*
- eiderdown: *a warm **cover** filled with feathers and put on top of the sheets and blankets on a bed A warm **cover** used without sheets or blankets is called a duvet*
- down: *in a **lower** place or at a **lower** level*
- a stab in the dark: *a guess or attempt which is not based on **knowledge** or experience and is very likely to be wrong or to fail*
- Advent calendar: *a **picture** usually of a Christmas scene with a series of 24 hidden **pictures** behind it Children open one part of the main **picture** to see one of the hidden **picture**s each day during Advent*
- turn the tables: *to succeed in gaining an **advantage** over someone who until now had an **advantage** over you*
- passage: *a **journey** or a ticket for a **journey** by ship to a place*
- guesswork: *the process of trying to find the **answer** to something by guessing or the **answer** found by using this method*
- revenge: *something that you do to **hurt** or punish someone because they have **hurt** you or someone else*
- unchallenging: *too easy and **therefore** not very interesting*