

is to provide the reader with a glossary of terms. This is particularly important for the many abbreviations that are used throughout. It would also be helpful for pinpointing small editing problems, such as the use of the abbreviation POS ('parts of speech') in Table 7, [p.31] while using COS (presumably 'categories of speech') in Table 10, [p.44] for discussing the same concept. A glossary would also help unite the disparate areas needed for understanding this material.

If one is inclined to think that this is an area that has been somewhat left untreated simply because of its relative unimportance, then a re-examination is clearly in order. Historical lack of attention to the most basic element of reading – sound to meaning correspondence – is an oversight in current dictionary design that should not be taken lightly. After all, if you cannot read or understand a definition, then why have the definition in the first place?

From a pedagogical standpoint, what *can* be learned from definitions? As Sobkowiak notes, incidental learning of vocabulary is well-known, but attention paid to learning from dictionary definitions is a rather neglected area of vocabulary-acquisition research: "During definition reading and processing by learners, incidental learning can occur, just like in any other reading activity...however [], I could find no research devoted to definition reading itself." [p.78]. If definitions can be improved so that sub-vocal reading is made easier (presumably leading to greater understanding of the definition), this would clearly be an improvement in dictionary development.

The book is a collection of several large-scale studies, compacted into one overall treatise. It provides a multitude of in-depth research programs that include:

1. An analysis of grapho-phonemic problems and inter-lingual phonological interference patterns encountered by Polish speakers learning English.
2. The development of a scale of the "Phonetic Difficulty Index" (PDI) – a coded metric of how difficult an English word would be for native Polish speakers to pronounce, based on the above analyses, and its application by algorithmic assignment to each entry of a reference wordlist database (a machine-readable version of the OALD wordlist).
3. Detailed general language and phonetic modeling, including an impressive array of statistical analyses, to act as baselines for comparing to dictionary-specific content.
4. Detailed empirical investigations of

the PDI metric, used for measuring the inherent phonologically-related difficulty of the following dictionary content:

- a. the defining vocabularies (DV) of four leading EFL dictionaries (LDOCE, OALD, CALD, and MEDAL)
- b. the definitions of the MEDAL dictionary
- c. 100-word samples of definitions from five EFL dictionaries (LDOCE, OALD, CALD, MEDAL, and COBUILD).

The basic findings indicate that these major dictionaries do not differ significantly from one another in terms of the PDI's of their defining vocabularies and definitions. Thus, no dictionary is 'phonetically harder' than any other. The question is, however, if some improvements could be made to make the dictionaries 'phonetically easier', and on what basis?

The comprehensive statistical analysis of MEDAL shows some differences in comparison to a reference lexicon. Some could be explained by the choice of DV or the usage of particular definition-specific words that boost the incidence of hard-to-pronounce phonemes. Sobkowiak points out that dictionary writers and editors could judiciously choose DV items or particular words in the definitions, keeping in mind the PDI metric. For example, the word 'whether', with the medial /th/ sound that is hard for Polish speakers to pronounce, could be replaced by the easier-to-pronounce 'if', while providing similar functionality in the definition (e.g. in the definition of *screen*: "to decide whether someone is suitable" vs. "to decide if someone is suitable", p. 90). Sobkowiak analyses dictionary microstructure and provides other such phonolapsological-based suggestions for making dictionary definitions easier for Polish learners.

Another finding is that the PDI metric, being word-based, does not capture across-word phenomena that are evident when words occur in various contexts. Having been involved in the application of phonetics to Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) for several decades, I can attest to the fact that this is not a trivial issue. Capturing contextually-variable coarticulation and vowel-reduction effects is a major obstacle in creating accurate acoustic models for speech recognition engines. Adequate across-word modeling, including intonation and other suprasegmental factors, is at the basis of providing natural sounding synthetic speech in TTS.

In the ELT sphere, it is apparent that 'a

## *Lexicography in Asia, Vol. 2*

The Department of English Language and Literature of the National University of Singapore and K Dictionaries are publishing a selection of papers from the ASIALEX 2005 Conference, as part of the *Lexicography in Asia* line. Entitled *Perspectives in Lexicography: Asia and Beyond*, it is edited by V.B.Y. Ooi, A. Pakir, I.S. Talib, P. Tan and I.J. Kernerman. The tentative contents are as follows:

**Introduction** · Vincent Ooi and Anne Pakir

### **Asian Perspectives**

- Lexical concerns about neologism in Chinese lexicography · Yihua Zhang
  - A three-way dictionary: bringing Cantonese, English and Putonghua together · Jacqueline Lam, Lan Li and Tom McArthur
  - Survey of pragmatic information in bilingual English-Chinese dictionaries · Saihong Li and Xiuying Wang
  - Grammatical class, tags and lemmas: a corpus-based study of the Malay lexicon · Gerry Knowles and Zuraidah Mohd Don
  - The treatment of greeting formulae in a bilingual dictionary of Gulf Arabic · Hashan Al-Ajmi and Lafi Al-Harbi
  - Words in Asian cultural contexts · Susan Butler
  - Dictionary making in an ascendant English knowing bilingual community · Anne Pakir
- Pedagogical Perspectives**
- Web-based EFL dictionaries: their development, characteristics, and problems · Shigeru Yamada
  - Exploring the potential of learner corpora for pedagogical lexicography · Yukio Tono

- Electronic dictionaries in the classroom · *Shinya Ozawa and James Ronald*
- The benefits of CD-ROM dictionaries in teaching · *Monika Szirmai*
- Designing a pedagogical dictionary of English academic vocabulary for Chinese speakers · *Gregory James*
- Dictionary use strategies by EFL learners in Taiwan · *Da-Fu Huang*
- English learners' dictionaries: an undervalued resource · *Julia Miller*
- Learner perspective for effective dictionary skill training and lexicography · *Tetsuro Fujii*
- Multiple word class entries in advanced learners dictionaries of English · *Sadayuki Nakane*
- Word combinations represented in illustrative examples in English learners' dictionaries · *Hai Xu*
- Not quite first language, not quite second language either: dictionary entries for learners caught in between · *Peter Tan*
- Competing lexicographical models and frameworks for the ESL context · *Vincent Ooi*

#### General Perspectives

- Interlingual lexicography, with special reference to research priorities · *Reinhard Hartmann*
- Reverse indexing and customization – future trends in bilingualized dictionaries · *Hugo T.Y. Tseng*
- Computer-mediated language: a new challenge for dictionary-making? · *Wengao Gong*
- Relational network notation and the intelligent web · *Jonathan Webster and Ian C. Chow*
- Linking dictionary and corpus · *Adam Kilgariff*
- Narrative and the multimodal contexts of lexicography · *Ismail Talib*
- Dictionaries and translation, from art to science? · *Ilan Kernerman*

word spoken in isolation' is only the beginning of pronunciation learning: the real test is if the word can be intelligibly spoken by the non-native in varying contexts, with proper stress, using varying intonation and dynamic syntactic patterning. As a former teacher-trainer involved in the technological hunt for the ultimate 'teacher-free' automatic program for teaching pronunciation to foreign learners, I can testify that this test is highly complex and reflects the intricate phonetic inter-dependencies that occur in the production of variable speech. The current international craze for accent reduction programs and the high attention paid to across-word contextual phenomena indicate that aiding such pronunciation problems addresses a real need; the success rate of such programs show that even human teachers (not only automatic-based instruction programs) find these difficult to adequately teach.

In this regard, Sobkowiak must be commended for his academic honesty in outlining such problems with the proposed PDI and the influence this may have had on the outcome of some of the results. However, nobody has yet produced a perfect metric the first time around, and this is where subsequent studies have their work cut out. It must also be noted that there are many possible applications for such a metric if it could be perfected, notably in the field of linguistic resource development for speech applications involving foreign accents, currently a pressing problem for ASR. Procedures for collecting databases that are relevant for speech recognition simply do not take into account difficulty of pronunciation. ASR databases typically contain recordings of hundreds, if not thousands, of speakers using prompt sheets that include linguistically designed material to cover phoneme variability related to contextual factors (e.g. the phonetically balanced sentences in the TIMIT database, Fisher et. al, 1986). Such collections of foreign speakers of English are difficult to create, since non-native speakers find it hard to read aloud the required material that must be recorded to create phoneme models (e.g. the 'Orientel' collection for several types of Arabic-accented English or French speech; Zitouni et. al 2002, Siemund et. al 2002).

Looking into foreseeable dictionary development, one can surmise that in the not-too-distant future it may be possible to have 'read-aloud' programs packaged into the dictionary itself, using real speech recordings or natural sounding TTS, for aiding the second-language

learner to read dictionary definitions. Until such time, however, users must still sub-vocalize, read, and understand these definitions. This research indicates that some type of 'phonetic control' can be accomplished to make the task easier, without impacting on other important lexicographic needs.

What can now be studied is the actual degradation of vocabulary learning that presumably would take place if very difficult phonetic material (based on the PDI metric) is used in the dictionary. Subsequent studies could model vocabulary learning difficulties based on PDI challenges, both within dictionary definitions and elsewhere, and, of course, for speakers of other languages learning English. It is hoped that in the future more attention will be paid to researching dictionary usage and effectiveness of definitions in terms of phonetic factors.

#### References

- Fisher W., G. Doddington, K. Goudie-Marshall. 1986. The DARPA Speech Recognition Research Database: Specifications and Status (TIMIT database), in *Proceedings of DARPA Workshop on Speech Recognition*. 93-99.
- Siemund R., D. Iskra, H. van den Heuvel, O. Gedge, S. Shammass. 2002. Multilingual Access to Interactive Communication Services for the Mediterranean and the Middle East, in *Specification of Validation Criteria*, ORIENTEL, Deliverable D6.2.
- Sobkowiak W. 2002. Phonetic keywords in learner's dictionaries, in U. Heid et al. (eds.), *Euralex 2000 Proceedings*. Stuttgart: IMS, 1.237-246.
- Sobkowiak W. 2003. Pronunciation in Macmillan English Dictionary for Advanced Learners on CD-Rom, *International Journal of Lexicography*, 16.4.423-441.
- Sobkowiak W. 2004. Phonetic keywords in EFL dictionaries revisited: MED, in M.C. Campoy and P. Safont (eds.), *Computer-Mediated Lexicography in the Foreign Language Learning Context*, *Colleccio Estudis Filologies*, 18. Castello: Universitat Jaume I.123-32.
- Zitouni I., J. Olive, D. Iskra, C. Choukri, O. Emam, O. Gedge, E. Maragoudakis, H. Tropft, A. Moreno, A. N. Rodriguez, B. Houft, R. Siemund. 2002. Orientel: Speech-Based Interactive Communication Applications for the Mediterranean and Middle East, *ICSLP-2002*.325-328.

**Shaunie Shammass**  
**K Dictionaries**  
 shaunie@kdictionaries.com