

EÖTVÖS LORÁND UNIVERSITY OF SCIENCES  
FACULTY OF HUMANITIES

DOCTORAL SCHOOL IN LINGUISTIC SCIENCES  
DOCTORAL PROGRAMME IN ENGLISH LINGUISTICS

PÉTER A. LÁZÁR

**PRINCIPLE AND COMPROMISE IN THE DICTIONARY**

INTERFACES OF THEORY AND APPLICATION  
IN LEXICOGRAPHY

DOCTORAL (PHD)  
DISSERTATION  
ABSTRACT

BUDAPEST 2010

## Main issues and structure of the dissertation

An exploration in lexicology and lexicography, the dissertation deals with the twin question of “what goes into the dictionary?”.

- (i) What entities that are supposed to be listemes in the *mental lexicon* are listed lexicographically?
- (ii) What are those *perceptions of linguistics* which find their way into, and will have a reflection in, dictionaries?

WHAT GOES INTO THE DICTIONARY is thus the title of the longest chapter, Chapter Two.

To comprehensively assess the insights of linguistics at large and summarize the nature of the mental lexicon (which lends itself to scrutiny less than other objects of study), a far larger body of knowledge would be needed than is available to this author. Concerning the nature of, and linkages between, the *listemes* of the mental lexicon, the literature enjoins great caution. The emphasis will thus be on the visible side – existing dictionaries as embodiments of this linguistic information, which can be lexicologically and lexicographically explored.

Lexicological *case studies* are presented to show how one depends on the other, and how much success in this endeavour may be hoped for. These include Chapter One: 1.6.1.1.1–2; Chapter Two: 2.4.6.2; 2.6.5.5.2, and most of Chapter Three, especially: 3.1.3; 3.1.3.1; 3.1.4.2.2–3; 3.1.5.2.2–3.

- 1 Particular instances of how *grammatical* information is presented lexicographically are examined, in case studies of Hungarian and English, and by scrutinizing entries of monolingual English as well as English↔Hungarian dictionaries. The assumption was that most advances of linguistic thinking that might be relevant are rather ill-suited to lexicographic representation, for one reason or other. A recurrent theme in the dissertation, this is in focus throughout Section 3.1, GRAMMAR IN THE DICTIONARY. For the purposes of this exploration a neat, if unnatural, separation of the grammatical and the lexical was assumed and the author’s doubt regarding their rigid duality temporarily suspended.
- 2 In any domain of linguistics, one will inevitably work with a number of claims usually couched in *dichotomies*. The dissertation seeks answers to the question whether lexicographic practice justifies these dualities, and whether they can be usefully employed in lexicography. A sketchy list of these dichotomies is given in 2.2. The assumption was that from a lexicographer’s angle, continua are better suited than discreteness to most phenomena, both in and outside lexicology. One such major duality is explored in 3.2, BETWEEN GRAMMAR AND LEXICON.
- 3 *Multiword expressions* (whose linguistic status is explored in 2.7–2.10) also point up the duality of “continua over crispness” forcefully. The lexicological literature offers no guidelines failsafe enough for use in dictionaries either to demarcate multiword expressions from other word combinations, or to adequately chart the internal structure of the extremely fuzzy category of multiwords (cf. 2.9.8).

- 4 Another theme throughout the study is *idiomaticity*, explored in some detail in the bulk of Chapter Two. The boundaries of this phenomenon, too, both externally and internally, are far too vague for consistent use in lexicography. The novelty of the dissertation is that it looks at the lexicographic repercussions of this fact through particular examples.
- 5 The dissertation proves one of its initial assumptions, that idiomaticity – in addition to being *gradient* both intra- and inter-categorially – covers a larger domain, with an application far wider, than is customarily assumed.
- 6 It is proved that the majority of (non-primary) compounds, the bulk of multiwords, and a great many hard-to-classify lexical items (e.g. binomials) are *idiomatic* in the sense of being *unpredictable*, which is argued to be the only true measure of idiomaticity. This is how compounds, “combining forms”, multiword expressions, and idiomaticity converge.
- 7 A theme throughout is the phenomenon of *productivity*. The dissertation proves, largely in accordance with the literature, that it traverses the domains of grammar and lexis. Productivity has been explored in the domain of compounds and multiwords, and its relation to different types of word – independent words, dependent words, semiwords, and bound bases – has been examined.
- 8 A cardinal claim of the dissertation is that lexicography should aim at *compromise between theory and practice*. This seems to have been corroborated by the case studies. It is thus suggested that it is better for lexicographic policies to be based on users’ needs than theoretical considerations, even where such information of a suitable kind is available. The contrasting of the lexicological case studies and the lexicographic practice as seen in existing entries, however, suggests that such information is rather conspicuous with its *unavailability*, and that even where it is available, further impediments on the users’ side may come into play, further widening the *gap between theory and user*.

## Conclusions and results

By contrasting lexicological and metalexigraphic ideals to existing dictionary entries, the dissertation has investigated interfaces of lexicography and lexicology, and lexicography and linguistics, to identify aspects where principled rigour can, or indeed must, be sacrificed to practical utility.

The dissertation substantiates the claim that *compromise* is to be sought between theory and practice. Lexicographic decisions are indeed more usefully made with users in mind than based on scraps of ill-fitting theory, even where such are available. This is in the spirit of Landau's admonition<sup>1</sup> that intelligibility must not be "sacrificed to a purity of style bordering on lunacy".

One finding, however, has been that such readily usable insights from linguistics are *not available*. If, however, most linguistic insights have little relevance for dictionary making (a gloomy view not universally shared), and/or they would be well nigh impossible to accommodate within lexicographic praxis (a form of pessimism that many analysts would subscribe to), and/or most of these insights would be lost on most users anyway (an almost consensual claim), then editors' familiarity with theory can at least help identify where those areas of compromise exactly are.

1. In charting the correspondences between "lists" in the mental lexicon vs. lists in dictionaries, only tentative hypotheses seem to be available as to the former. Based on the little that is available, the dissertation has proved that the printed lists produced by lexicographical wisdom are far removed indeed from the "lists" of the mental lexicon. *Next to nothing* that is assumed, however tentatively, about the mental lexicon can be *reflected by means of lexicography*.

2. The dissertation assumes that the mental lexicon is a system of *multiword units rather than a set of isolated words*. Unfortunately for dictionary-making, however, these are so varied and elusive, and as a consequence, their linguistic treatment is so chaotic, as to promise very little for practitioners of lexicography. A novelty of the dissertation is a lexico-semantic and grammatical examination of such lexical units of diverse types above the traditional word level, and contrasting these with potential and existing dictionary entries.

3. The dissertation claims, and proves on a variety of examples, that very few findings of linguistics have made their way into lexicography; to show the little that has been utilized, entries from several dictionaries of different types have been explored. More recent dictionaries that have been under scrutiny include: RHWUD 1999 (*Random House Webster's Unabridged Dictionary*); MWUD 2000 (*Merriam-Webster's Unabridged Dictionary*); CC 2003 (*Collins COBUILD on CD-ROM*); AHD 2004 (*American Heritage® Dictionary of the English Language*); LDCE 2005 (*Longman Dictionary of Contemporary English*); OALD 2005 (*Oxford Advanced Learner's Dictionary*); MED 2002 & 2007 (*Macmillan English Dictionary*); CALD 2003 & 2008 (*Cambridge Advanced Learner's Dictionary*); CCAD 2009 (*Collins COBUILD Advanced Dictionary*).

4. Several claims of the dissertation are best couched in terms of – not just lexicographic and lexicological – *paradoxes*. It has been found and illustrated throughout that notions of *gradience* or *cline* are better suited to many phenomena than *discreteness*, both in and outside of lexicology. This, the present author thinks but has not argued here, probably carries over to linguistics at large. (By the recognition of prototypicality and the indeterminacy of linguistic facts, the dissertation does

---

<sup>1</sup> Landau (2001) .

not automatically argue for prototypes in the realm of things “out there”, and thus assumes no specifically cognitive framework.)

4.1 Perhaps the *gravest paradox* of lexicography is this: almost nothing of the *analogue* nature of the lexicon, and possibly language at large, can be represented in the inevitably black-and-white, two-dimensional world of lexicography, which, in addition to this crispness, must add a good measure of inevitable simplification for the sake of users. This amounts to the claim that although in lexical phenomena the probabilistic nature of language clearly transpires, lexical patterns are difficult to capture in the static world of dictionaries. Also, it seems that what users expect are exactly hard-and-fast rules and rigid boundaries supplied by dictionaries that change as little as possible.

4.2 *Idiomacity* itself has been found to be *gradational*, especially that of *multiword expressions*, which show a strong centre–periphery pattern in other respects as well. I have tried to show that this adds to the difficulty of their lexicographic treatment, whose inadequacy is quite clear in the most recent monolingual native speaker dictionaries, and less so in the didactically otherwise successful learner’s dictionaries. This edge of learner’s over native speaker dictionaries is a paradox specifically in English lexicography: the mundane considerations and motives of the ELT industry have apparently left on it a deeper impact than has theoretical linguistics.

4.3 The dissertation has separated the notions of *idiomaticity*, *transparency*, and *compositionality*<sup>2</sup>. Idiomaticity is best seen as defined in terms of *unpredictability*. In that framework, all multiwords – including compounds, combining forms, binomials, and what are referred to in the dissertation as traditional “colourful” idioms, are idiomatic since they are unpredictable across any two languages, which is the *only true measure of idiomaticity*.

4.4 One type of gradience is inherent in the notion of *productivity*: the status of lexical units ranging from frequent/existing through rare/existing and potential all the way to the ungrammatical but existing. The dissertation (especially in 3.2) captures another paradox of dictionary-making: that productivity is also impossible to capture in the static world of dictionaries, where, due to the inevitable limitations, “all words are equal”.

4.5 Another paradox identified in the dissertation is that while not just idiomaticity but *many (perhaps most) lexical phenomena inhabit continua*, this ill suits the purposes of lexicography. Dictionaries, especially “instrument dictionaries”, where *clear-cut advice* is expected, produce *tools* for the layperson. Aggravating this problem are the efforts at *simplification*, motivated by users’ perceived needs and general poor levels of grammatical informedness.

5. Numerous examples support the claim that linguistic “theory”, more and more aloof with its erratic changes and spectacular fragmentation, proves less and less adequate for a provision of guidelines for lexicographic practice, especially as regards the narrower field of *grammar*. It seems evident that *most linguistic advances are lost on lexicography*. At the linguistic input is a collection of arcane, mostly incompatible and ever-shifting findings never meant for lexicographic end use in the first place. At the output, in the dictionaries, which under user *pressure to change as little as possible* anyway, is a hardly noticeable imprint of linguistic science, with the gap between it and lexicography ever widening.

6. If users’ needs rather than theoretical considerations can be the real measure of evaluating works of lexicography, then the inadequacy of linguistic theory as model or theoretical basis, serving as a stable source of inspiration, is *less damaging* than might be supposed. Users’ needs, however, are much harder to assess than is usually thought. There are too many different users, and indeed too widely differing tasks in any bilingual relation. And while the *need for user research* is

---

<sup>2</sup> Drawing on Nunberg & al. (1994).

overemphasized in the literature, what this injunction precisely means for the praxis is unclear. Moreover, the “know your user” requirement, unfortunately, is unhelpful for small markets, and especially so for bilingual Hungarian↔English lexicography, where dictionaries are produced for Hungarians using English rather than vice versa, and thus will never have a chance of being as diverse as to justify large-scale user studies.

This means that even if users’ needs can indeed be found out, dictionaries cannot be easily tailored according to the findings. This suggests that although dictionaries will evidently be different, very few, if any, of their differences in design and editorial policy will come from user/market research. The dissertation arrives at the indeed unfortunate conclusion that because input from linguistics is insufficient and user research sparse, perhaps inevitably dictionaries will be bound to continue on their path of inertia.

## Potentialities for further research

The dissertation, “a lexicographer’s view of linguistics”, has brought to the surface quite a few further issues which are of paramount importance. A most worthwhile path of further study is to explore and chart, in a consistent and unified framework<sup>3</sup>, the *gradience* of various types of *lexical* phenomena.

A list of sources pointing to further areas of exploration and thus relevant have become available after closing the manuscript is provided below. One of them<sup>4</sup>, a study that claims that “the standard word class framework becomes just a convenient labelling system, primarily for such everyday purposes as teaching as well as dictionary and grammar writing but without any theoretical background”, is thus a pointed illustration of an even more radical rift between linguistic theory and lexicography.

The dissertation refers to Bauer (2003)<sup>5</sup>, according to which some processes are more while some less (centrally) morphological and more syntactic than others, and thus may be argued to exhibit a continuum of major significance. Prefixation has affinities with (neo-classical) compounding as well as back-formation; suffixation is borders on neo-classical compounding, conversion, and back-formation; the latter shows similarities with both conversion and clipping; neo-classical compounding has affinities with blending, which is similar to acronym formation. Crucially, compounding borders on syntax. Within the theory, this may necessitate redrawing on a major scale of the boundaries of morphology and consequently of the compartments of language. The aim of lexicographic practice, by contrast, will be to *capture as many as possible* and represent as many as can be *usefully represented*, of these potential changes.

---

<sup>3</sup> In the spirit of Aarts (2007).

<sup>4</sup> Kenesei (2010); my translation.

<sup>5</sup> Bauer (2003:122–125).

## Sources relevant for further research only recently available

- Booij (2009): Geert Booij “Compound construction: schemas or analogy? A construction morphology perspective”. To appear in: Sergio Scalise & Irene Vogel eds. *Compounding*. Benjamins.
- Geeraerts (2010): Dirk Geeraerts *Theories of lexical semantics*. OUP.
- Kastovsky (2009): Dieter Kastovsky “Astronaut, astrology, astrophysics: About Combining Forms, Classical Compounds and Affixoids”. In: R. W. McConchie, A. Honkapohja & J. Tyrkkö eds. *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis*, 1–13. Somerville.
- Kenesei (2010): “Vannak-e szófajok? És ha igen, mennyi?” [Whether word classes exist, and if yes, how many?] Presentation at the Research institute for Linguistics of the Hungarian Academy of Sciences, 17/06/2010, <http://www.nytud.hu/kenesei/szofajokea100617.pdf>, accessed 27/06/2010.

## Most important literature in the dissertation

- Aarts (2007): Bas Aarts *Syntactic Gradience. The Nature of Grammatical Indeterminacy*. OUP.
- Aarts & McMahon (2007): Bas Aarts & April McMahon eds. *The handbook of English linguistics*. Blackwell.
- Ameka (2006): F. K. Ameka “Interjections”. In: Brown (2006)
- Atkins & Rundell (2008): B. T. Sue Atkins & Michael Rundell *The Oxford Guide to practical lexicography*. OUP.
- Ayto (2006): John Ayto “Idioms”. In: Keith Brown (2006).
- Bauer (2001): Laurie Bauer *Morphological productivity*. CUP.
- Bauer (2007): “Compounds and minor word-formation types”. In: Aarts & McMahon (2007)
- Béjoint (2000): Henri Béjoint *Modern lexicography: An introduction*. OUP.
- Booij (2007): Geert Booij *The grammar of words. An introduction to morphology*. 2nd ed. OUP.
- Brown (2006): Keith Brown ed. *Encyclopedia of Language and Linguistics*. Elsevier.
- Cowie (2001): Anthony P. Cowie ed. *Phraseology: Theory, Analysis, and Applications*. OUP.
- Fontenelle (2008): Thierry Fontenelle ed. *Practical lexicography. A reader*. OUP.
- ten Hacken (2009): Pius ten Hacken “What is a dictionary? A view from Chomskyan linguistics.” In: *International Journal of Lexicography*, Vol. 22 No. 4. OUP.
- Hanks (2006): Patrick Hanks “Lexicography: Overview”. In: Brown (2006)
- Hudson (1988): Richard Hudson “The linguistic foundations for lexical research and dictionary design”. In: *International Journal of Lexicography*, Vol. 1 No. 4. OUP.
- Jackson (2002): Howard Jackson *Lexicography. An introduction*. Routledge.
- Kenesei (2007): I. Kenesei “Semiwords and affixoids. The territory between word and affix”. In: *Acta Linguistica Hungarica* 54: 263-293.
- Kiefer (1990): Ferenc Kiefer “Linguistic, conceptual and encyclopedic knowledge: some implications for lexicography”. In: T. Magay & J. Zsigány eds. *BudaLEX '88 proceedings. Papers from the 3rd International EURALEX Congress Budapest, 4–9 September 1988*.
- Landau (2001): S. I. Landau *Dictionaries. The art and craft of lexicography*. 2nd ed. CUP.
- Moon (1998): Rosamund Moon *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford Studies in Lexicography and Lexicology. Oxford: Clarendon Press
- Nunberg & al. (1994): Geoffrey Nunberg, Ivan Sag & Thomas Wasow “Idioms”. In: *Language*, 70:3.
- Országh (1967): László Országh “A plea for a dictionary of modern idiomatic English”. In: Virágos (2007). [originally in Volume III of *Hungarian Studies in English* 1967, 71–81]
- Plag (2006): Ingo Plag “Productivity”. In: Brown (2006)
- Stubbs (2002): Michael Stubbs *Words and phrases. Corpus studies of lexical semantics*. Blackwell.
- Wierzbicka (1993): Anna Wierzbicka “What are the uses of theoretical lexicography?” In: *Dictionaries: The Journal of the Dictionary Society of North America* 14: 44–78.
- Wray (2006): Alison Wray “Formulaic language”. In: Brown (2006).

## Related publications in the subject of the dissertation

- “Degree words and lexical functions”. In: *English in function*, ELTE, Budapest, 1986.
- “Review of Thompson’s Dictionary for Hungarian Learners of English”. In: *IATEFL Network-H Newsletter*, Budapest, 1991.
- “A Cambridge dictionary”. In: *Nyelvinfó* Vol. 3, No. 3, Budapest, 1995.
- “Cambridge International Dictionary of English”. In: *Modern nyelvoktatás*, Vol. 1, No. 1, Budapest, 1995.
- “Oxford Dictionary of Computing for learners of English”. In: *Modern nyelvoktatás*, vol. 2, No. 4, Budapest, 1996.
- “Breaking new ground: the Cambridge International Dictionary of English”. In: *The Even Yearbook 2, ELTE SEAS Working Papers in Linguistics*, Budapest, 1996.
- “WordSmith Tools” In *Modern nyelvoktatás* Vol. 3, No. 4, Budapest, 1997.
- “Key word in context: teaching syntax through concordances”. In: *Proceedings of the 1996 EUROCALL Conference*, Szombathely, 1997.
- “Towards a taxonomy of false friends: Faux amis for Hungarian users of English”. In: *The Even Yearbook 3, ELTE SEAS Working Papers in Linguistics*, Budapest, 1998.
- “The kind of dictionary you need but aren’t quite sure exists” (with György Varga). In: *NovELTy* Vol. 6, No. 1, British Council – IATEFL-H Budapest, 1999.
- “Milyen szótárra van szüksége a felhasználónak?” (with György Varga). In: *Kultúra és közösség*. Vol. 3, No. 3–4, Budapest, 1999.
- “Kurrens szócskák, friss idiomák”. Reviews of two Cambridge dictionaries. In: *Tandem* Vol. 4, No. 4, Budapest, 1999.
- “Oxford Advanced Learner’s Dictionary”. In: *Modern nyelvoktatás* Vol. 6, No. 2/3, Budapest, 2000.
- Angol–magyar szótár* (with György Varga), Aquila, 2000.
- Magyar–angol szótár* (with György Varga), Aquila, 2000.
- Angol–magyar kéziszótár* (with György Varga), Aquila, 2001.
- Magyar–angol kéziszótár* (with György Varga) Aquila, 2001.
- “PASSWORD – English–Hungarian Learner’s Dictionary”. In: *Modern nyelvoktatás* Vol. 9, No. 4, Budapest, 2003.
- “Kétnyelvű CD-ROM nagyszótárak (Akadémiai Publishers)”. In: *Modern nyelvoktatás* Vol. 10, No. 2–3, Budapest, 2004.
- Angol–magyar szótár* 2nd edn. (with György Varga) Aquila, 2005.
- Magyar–angol szótár* 2nd edn. (with György Varga) Aquila, 2005.
- “What is in a dictionary? – Lexicography”. In: Richter Borbála ed. *First steps in theoretical and applied linguistics*, Budapest, 2006.
- “Lexical semantics – Words, meanings, and their relationships”. In: Richter Borbála ed. *First steps in theoretical and applied linguistics*, 2006.
- “A szótár mint lista. Listázottság a mentális és a nyomtatott szótárban.” In: Tibor Frank & Krisztina Károly eds. *Anglisztika és amerikanisztika. Magyar kutatások az ezredfordulón*. Tinta, Budapest, 2009.
- Angol–magyar szótár* (with György Varga) Pro-Team, 2010.
- Magyar–angol szótár* (with György Varga) Pro-Team, 2010.