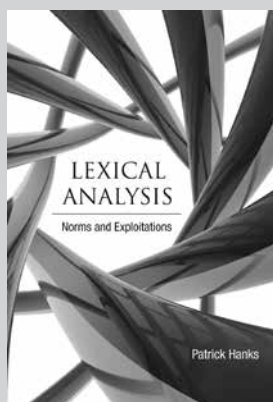


## Patrick Hanks. *Lexical Analysis: Norms and Exploitations*



**Lexical Analysis**  
**Norms and Exploitations**  
 Patrick Hanks  
 Cambridge MA: MIT Press  
 2013. 478 pages  
 ISBN 9780262018579

This book is a long-awaited milestone of lexicology, lexicography, and empirical theories of language. As the title suggests, its main focus is on the analytic task that is central to lexicography, but the scope of the work inevitably reaches into other aspects of the broader lexicographical undertaking. Ultimately, the book and its theory set a course toward uncharted territory, where lexical analysts will create language resources that may bear little resemblance to dictionaries as we know them. But the book is not only about lexicography: it aims to establish a complete evidence-driven theory of language.

### Overview

*Lexical Analysis: Norms and Exploitations* (henceforth LA:N&E) unifies Patrick Hanks's work on the Theory of Norms and Exploitations (henceforth TNE) in thirteen chapters, of which four have been published elsewhere over the past two decades. Readers who have read the previous publications should avoid merely skimming them in this edition. Much as words have different meanings in different contexts, the new context of this book brings to the earlier papers meaning and nuance that is not as apparent when they are read in isolation.

Chapters 1–3 establish basic concepts and theoretical foundations: why do we need a new approach to meaning? What are 'words', and what exactly might 'meaning' consist of? Most of us will have read these kinds of discussions many times before, but these chapters are where Hanks defines the assumptions and terminology that drive the rest of the book, so they are an important preamble to the work in its particulars.

Chapters 4–8 are the meat of the theory. Chapters 4–6 describe prototypes, patterns, and norms & exploitations, and firmly assert the central importance of context to understanding lexical items. Chapter 7 introduces Hanks's detailed account of alternations, which are variations in pattern structure that do not change the normal meaning of a clause's components. Chapter 8 discusses exploitations in broad scope and detail, finally presenting the first revision of a typology of exploitations. These two chapters broaden the footprint of the theory enough to make it a very comfortable space for further research.

Chapters 9–13 explore applications of TNE to problems that range far afield of lexicography. Chapter 9 reaches into

literature to exemplify how creative writers are creative exploiters of word meaning, and to consider the ramifications of that creativity on the process of lexical analysis and corpus creation. Chapter 10 discusses the range of complexity that is possible in different words' meanings (or meaning potentials): some can be very complex, but not all types of words have the same level of complexity, and so different levels of analysis are appropriate for the different types of words Hanks accounts for.

Chapters 11 and 12 are surveys of preceding approaches to meaning and the lexicon, evaluated in light of TNE. Chapter 11 traces western philosophies of meaning from Aristotle to the present, while 12 considers treatment of the lexicon by previous linguistic theories. Confronting other theories of meaning and lexicon gives Hanks a wealth of opportunities both to credit previous scholarship and to assert how TNE differs from other theories. These chapters are not dutiful surveys of related research: they are dynamic debates with the theorists who have defined our common understanding of words and meaning.

Given the diverse backgrounds of working lexicographers, it is quite possible that many professionals will not have studied the majority of the scholarship that Hanks confronts. The contextualization that Hanks gives these theories is inevitably prejudicial in Hanks's favor, but these chapters serve to put the daily practitioner on equal theoretical footing with the well-read academic.

Chapter 13 reviews the "broader picture", adds references to late-breaking research, considers implications for other fields such as language teaching, and points to many areas for further development of the theory. The book is remarkably strong on concluding material overall. Most of the chapters conclude with sections of "Implications" and a bullet-pointed "Summary", both of which reiterate key arguments and concepts. These sections are a nice feature that more books should have.

The central theme of the book is the usefulness of "distinguishing systematically between normal patterns of collocations and creative uses of those patterns" (p 6). If analysts make such distinctions, and then set aside the exploitations and focus on the norms, Hanks argues that "some well-known linguistic problems are largely solved" (p 17). Hanks hopes that TNE will resolve enough problems that it may be

applied as an evidence-driven theory of language as a whole, rather than simply a theory of lexis and semantics.

The nature of linguistic evidence drives several other core themes. Hanks argues, with passion and wit, against the use of invented examples in any part of empirical language work. He acknowledges the inescapable fuzziness of linguistic events, and appeals to the central role of context in reducing this fuzziness and creating ‘meaning’. He accounts for ‘meaning’ as a confluence of fuzzy ‘meaning potentials’ rather than as a selection from among delineated, mutually-exclusive options. The tension between empiricism and intuition is a standard theme of works in corpus-driven lexical analysis, and Hanks shows that statistical analysis is not enough to overcome the biases and prejudgments of intuition. To accompany statistical measures, he offers many rules and heuristics for deciding whether a given linguistic event belongs in one fuzzy category or another.

#### Hurdles for the reader

The notion of “setting aside the exploitations” appears to be a common stumbling block for lexicographers upon their first exposure to TNE (at least, it was for this reviewer upon my first reading of Hanks’s 1994 paper). When Hanks tells you that a given word has five “normal” meanings, your creative brain immediately remembers sixth and seventh meanings that you saw written by ‘reputable’ authors just the other day. A lexicographer who has been trained to account for all uses of a given definiendum may feel justifiably uncomfortable if told to pass over a frequent collocation that has a discrete meaning.

In answer to such concerns, Hanks makes compelling arguments that the distinction between normal and creative usage is essential, and can indeed be made empirically, even if allowing for some fuzziness at the edges. Hanks makes a distinction between cognitive salience (“memorable because it is unusual”) and social salience (“conventional and for that reason unmemorable”) (p 5). This distinction, together with chapter 9’s survey of literary exploitations, build a strong case that the tremendous variety of possible exploitations can be managed only by means of a solid understanding of what happens in normal usage.

Hanks lays out clear and principled criteria to tell the difference between norms, exploitations, alternations, and simple errors. These criteria provide coherent theoretical justification to some analytic conclusions which, heretofore, even careful

analysts of corpus evidence may reach only ad hoc or through intuition.

Of course, even the clearest criteria may not always prevail over unclear data, but a satisfying majority of cases appear to be resolvable with the techniques that Hanks has refined and developed in the course of his career. Furthermore, it is not the case that fuzziness creates usage that is utterly incomprehensible to humans or uncategorizable by machines. Rather, part of the strength of the theory is that it leverages normal uses to make sense—as much sense as can be made—of all the varieties of non-normal usage, in ways that show good signs of being computationally tractable.

Readers not familiar with frame semantics (Fillmore 1982), generative lexicon theory (Pustejovsky 1995), or Hanks’s earlier work around TNE will, in a few places, need to be indulgent of certain formal technical description that is foreshadowed or introduced without being fully explained: I refer here chiefly to the double-square-bracket markup of the corpus patterns, which begins to appear in sprinkles in chapter 2, but is most fully explained only just before it flows in a torrent in chapter 4. A page or so of chapter 2 seemed to me to be more appropriate for a later chapter, but a patient reader will benefit from the warm-up if they take the time to grok it in chapter 2.

#### Hurdles for the theory

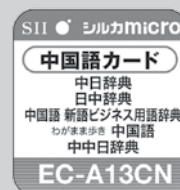
Hanks’s productivity and influence means that the TNE has been well-exercised, and in many places it reflects consensus among the schools of corpus analysis and lexicology of which Hanks is a major figure. Unfortunately, in lexicography it is an acknowledged truth that much of the theoretical consensus is not implemented within the most widely consumed lexical references. This may delay the impact of the theory on lexicography, and in the absence of lexical resources built on these principles, computational linguistics will not be able to reap the theory’s full benefit without first helping to create those resources.

Hanks has been at work to remedy the gap between theory and practice, with the Corpus Pattern Analysis project and subsequent *Pattern Dictionary of English Verbs*. These projects are an exciting, practical testing ground for the theory, and Hanks’s eminence in practical lexicography emphasizes the applicability of the theoretical methods he proposes here. Considering the troubled state of the dictionary business as we know it, it is possible that a complete lexicon on TNE principles may not be made by a “dictionary” group. Still, the book’s tips,

#### KERNERMAN JAPANESE



Series of bilingual learners’ dictionaries for speakers of Japanese on SD cards for MICRO hand-held e-dictionaries by Seiko Instruments, Tokyo 2013.



#### Chinese 中中日辞典

Chinese: Saihong Li (chief editor), Huimin Li, Xilan Sun, Yuan Yuan Wang, Chao Zhang, Hao Chen, Chia Chen Liu

Japanese translation: Koji Kawa



#### French 仏和辞典

French: Nathalie Lanckriet (chief editor) | Valérie Lecœur, Marine Lesprit, Frédérique Melanie-Becquet, Emilienne Morruzzi, Caroline Moulart-Fremaux, Edouard Trouilleux, Colette Alouch

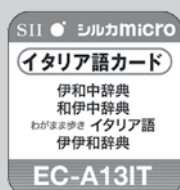
Japanese translation: Ken’ichi Hisamatsu, Fumitoshi Hayakawa, Yuji Obataya



#### German 独独辞典

German: Susanne Billes (chief editor), Monika Böhme, Sabine Corso, Susanne Dyka, Marion Hahn, Christine Moehrs, Emmanuelle Worbs

Japanese translation: Yoshinori Kaisaki, Naoko Toshimitsu, Tamaki Sakurada-Kaufmann



#### Italian 伊伊和辞典

Italian: Palma Gallana (chief editor), Davide Levi  
 Japanese translation: Yu Akao, Castrese Cacciapuorti, Yoko Shibue



#### Korean 韓韓日辞典

Korean: Youn-Sill Couralet (chief editor), Stéphane Couralet, Migyung Seok, Kyung-Ryang Kim, Kwi-Jeong Yoon  
 Japanese translation: Shigeru Hashimoto



#### Russian 露露和辞典

Russian: Margarita Sapozhnikov (chief editor), Lana Philippov, Polina Brukman, Hanna Ioffe  
 Japanese translation: Nobuyoshi Adegawa, Elena Baibikov, Yuka Ikeda, Tomoko Masuda, Yuko Wakabayashi



#### Spanish 西西和辞典

Spanish: Miguel Eduardo Montoro (co-chief editor), Irene Renau Araque (co-chief editor), Sergio Mario López, María Verónica Elizondo, Carlos Sebastián Ceró  
 Japanese translation: Kobayashi Kazuhiro, Shingo Kato, Keisuke Gozawa

[http://www.sii.co.jp/cp/products/card/index\\_micro.html](http://www.sii.co.jp/cp/products/card/index_micro.html)

insights, and simple rules of thumb are as actionable for anyone working with corpora as they are for lexicographers.

Among the obiter dicta that pepper the book, there is mention that “expectation of homogeneity” is a trap for the “unwary [lexical] analyst”—a warning that “there is no reason to believe that the apparatus required for the semantic analysis of verbs will be necessary or sufficient for the semantic analysis of nouns (or vice versa)” (p 16). Hanks describes an apparatus for verbs and also for nouns, but spends much more time on the verbs. Indeed, Hanks’s current project is a pattern dictionary of 3,000 English verbs. It would be interesting to know how the book and the theory might have developed differently if Hanks’s ongoing projects focused on nouns instead of verbs.

It also seems relevant for a complete theory of language to consider the interactions between patterns of nouns and of verbs. LA:N&E is also virtually silent on the other parts of speech, aside from mentions that the TNE is driven by content words. Hanks acknowledges that the theory is incomplete, and in many places points to further work that is necessary. To paraphrase Melville, small theories may be finished by their first architects; great ones ever leave the coperstone to posterity. Recently, Ken Litkowski has undertaken a Corpus Pattern Analysis project for prepositions (Litkowski 2012) that may precipitate some answers to how patterns interact.

Early in the book, Hanks notes that ontologies have yielded disappointing results for word-sense disambiguation, and attributes this to severe underestimation of the fuzziness of boundaries between categories. TNE offers a very different model of meaning than what is traditionally considered in word sense disambiguation (WSD), and does not waste energy trying to avoid fuzziness, instead quarantining boundary cases from normal behavior.

Far larger-scale implementations are necessary to learn whether the theory can reduce the underestimation enough to serve a purpose like sense disambiguation (using whatever empirically-derived substitute for WSD is most applicable, since nobody who seriously works with meaning believes in ‘word sense’ as such (Kilgarriff 1997)).

In addition, LA:N&E does not address the question of how reproducible Hanks’s own analyses might be. In reading the book I occasionally found myself disagreeing not with the broader theory, but with individual judgments that Hanks had made in describing the theory’s practical output. No matter how evidence-driven the theory is, human analysts will not always interpret the

same evidence in the same way. Fuzziness is one thing, but undoubtedly some usages will defy agreement. Will TNE be able to yield reliable inter-annotator agreement on the whole?

#### The future that the TNE points to

LA:N&E is only the bottom of the bottom-up theory of language that Hanks envisions. The book points to many areas where the theory needs further development, and leaves many questions unanswered. How well does TNE apply to languages other than English? Do other languages have types of exploitations that are not possible in English? How are word senses activated by context? Without clear criteria, “sense activation” will remain merely a hiding-place for intuition.

A very real question is whether this kind of analysis still needs to be done by humans. Hanks does not extensively address the prospect of doing lexical analysis computationally—the theory is expressed for the human analyst. Some have seen this as under-ambitious, supposing that if we can do it with our brains, we should be training machines to do it. I believe that we can train machines to “do” TNE, but we cannot train them without first understanding what exactly the task is. For both human and computational implementations of this theory, Silvie Cinková (and her group at Charles University in Prague) have begun investigating the capacity for humans to be trained in Corpus Pattern Analysis (Cinková 2012a) as well as ways to manage the fuzziness of semantic categories (Cinková 2012b). Hanks also points to Popescu (2012), as a path to automating the processing of corpus patterns in the not-too-distant future. These efforts may ultimately be what closes the gap between theory and practice, both for human and computational lexicography.

#### The value of the TNE

This is not a book of abstract theory that cannot be practically implemented. The methods described here can be used by anyone who works with corpus evidence, on a large or small scale. A work with the aims of the *Pattern Dictionary of English Verbs* is a major undertaking, but even for the daily practice of lexicography on existing projects, the book offers many simple heuristics for the classification and definition of corpus evidence, which can be used as soon as the principles are understood. Indeed, once the principles are understood, any other way of handling corpus evidence may seem unacceptable.

The book and the theory aim to do much more than improve lexicography. Working

through all the details of the empirical basis of TNE yields a theory and a practical framework that shows strong potential to help with many of the other difficult problems in linguistics. The book is overflowing with suggestions for future research, any of which will solidify the undeveloped parts of TNE while also advancing the sciences of language. LA:N&E is an important work, a milestone of several fields, but there is still much work to do.

#### References

- Cinková, S., Holub, M., Kríž, V. 2012a.** Optimizing semantic granularity for NLP - report on a lexicographic experiment. In *Proceedings of the 15th EURALEX International Congress 2012*, 523-531, European Association for Lexicography, Oslo.
- Cinková, S., Holub, M., Rambousek, A., Smejkalová, L. 2012b.** A database of semantic clusters of verb usages. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 3176-3183,

European Language Resources Association, İstanbul, Turkey.

- Fillmore, C. J. 1982.** Frame Semantics. In *The Linguistic Society of Korea* (ed.), *Linguistics in the Morning Calm*: 111-137. Seoul: Hanshin Publishing Co.
- Hanks, P. 1994.** Linguistic Norms and Pragmatic Exploitations, or Why Lexicographers need Prototype Theory, and Vice Versa. *Papers in Computational Lexicography: COMPLEX '94*, 89-113.
- Kilgarriff, A. 1999.** I don't believe in word senses. *Computers and the Humanities*, Vol. 31, No. 2, 91-113.
- Litkowski, K. 2012.** Corpus Pattern Analysis of Prepositions, Technical Report 12-02. Damascus, MD: CL Research.
- Popescu, O. 2012.** Building a resource of patterns using semantic types. In *Proceedings of LREC 2012, Istanbul*.
- Pustejovsky, J. 1995.** *The Generative Lexicon*. Cambridge, MA: MIT Press.

**Orion Montoya**  
orion@mdcclv.com



#### KERNERMAN Semi-Trilingual English-Korean-L3

Afrikaans, Arabic, Bulgarian, Chinese Simplified, Chinese Traditional, Danish, Dutch, French, German, Hindi, Hungarian, Indonesian, Italian, Japanese, Malay, Norwegian, Polish, Portuguese Brazil, Russian, Serbian, Spanish, Swedish, Thai, Turkish, Ukrainian, Urdu, Vietnamese

K-Apps for Android  
Daolsoft, Seoul. January 2013

## Launching of *LEXICOGRAPHY: Journal of ASIALEX*

The long-awaited, flagship publication of the Asian Association for Lexicography will be launched in 2014. *LEXICOGRAPHY: Journal of ASIALEX* will be published by Springer, the prestigious worldwide scientific and professional publisher. It will have two issues a year, appearing in April and October, in print and online. The chief editor for the initial three-year term is Professor Yukio Tono, of the Graduate School of Tokyo University of Foreign Studies and current President of ASIALEX, and the editorial board consists of Gilles-Maurice de Schryver (Ghent University), Chu-Ren Huang (Hong Kong Polytechnic University), Shin Ishikawa (Kobe University), Ilan Kernerman (K Dictionaries), Deny A. Kwary (Airlangga University), Lan Li (Hong Kong Polytechnic University) and Shigeru Yamada (Waseda University), with the assistance of an advisory board including international scholars.

ASIALEX was established in 1997, with the principal mission of fostering scholarly and professional activities in the fields of lexicography and dictionaries in and for Asia. There are time-honored lexicographic traditions in Asia and great many topics to discuss and problems to solve. However, the new journal aims to serve as a leading-edge crossroad and powerhouse for all global issues of lexicographic interest, with an emphasis on Asian perspectives and concerns. Just like the ASIALEX conferences, *LEXICOGRAPHY* is open for the world's researchers, lexicographers, dictionary makers,

translators, teachers, students, and all language lovers to discuss issues concerning lexicography and dictionary history, typology, use, criticism, structure, components, compilation, application, media, and interaction with other fields such as phraseology, translation, corpus linguistics, language learning, ICT, etc. Subscription will entail membership to ASIALEX and the right of free publication to authors, subject to peer review of papers.

The text above is based on the ASIALEX announcement that was drafted by Shigeru Yamada (with additions by Chu-Ren Huang and Deny A. Kwary) and revised by me. I had the privilege to negotiate on behalf of ASIALEX the framework of this publication and cooperation with Springer for the last year, and would like to thank my colleagues on the board for their support and Springer's representative, Federica Corradi Dell'Acqua, for her dedication. During the recent DSNA meeting at the University of Georgia in May 2013, I had the opportunity to discuss our forthcoming publication with the editors of *Dictionaries* and *International Journal of Lexicography*, respectively Edward Finegan and Anne Dykstra, who heartily welcomed the news. We look forward to establishing *LEXICOGRAPHY* as a prominent international journal and proceeding to collaborate with all our sister publications around the world.

**Ilan Kernerman**