The Role of E-lexicography in the Confirmation of Lexicography as an Independent and Multidisciplinary Field

Theo J.D. Bothma, Rufus H. Gouws, Danie J. Prinsloo

University of Pretoria, University of Stellenbosch, University of Pretoria
e-mail: theo.bothma@up.ac.za, rhg@sun.ac.za, danie.prinsloo@up.ac.za

Abstract

This paper argues in favour of an adapted, extended theory of lexicography to cater for the fast-growing e-lexicography practice. We believe that such a modern general lexicographic theory should be rooted in three fundamental phases of the development of lexicographic theory, i.e. the focus on linguistic contents – Zgusta, Wiegand's focus on dictionary structures and Bergenholtz and Tarp's emphasis on lexicographic functions with the user in focus. Although the linguistic contents and the functions of printed and e-dictionaries can be similar, the presentation differs and that demands innovative formulations of dictionary structures. Lexicographic theory is expanded by describing and taking cognisance of all innovative technological features enabled by the computer era, including data collection and processing and data representation and interface design, in order to respond to the different needs of the envisaged target users in an optimal way. Insights from the field of information science facilitate the compilation of a much more sophisticated specification of user needs and the way in which users interact with the information. The lexicographer has an obligation to provide users with guidance and required information whilst guarding them against information overload.

Keywords: lexicographic theory; interdisciplinary; multidisciplinary; information science; user support tools

1   Introduction

Since the introduction of lexicographic theory, the focus shifted from an emphasis on the linguistic contents of dictionaries to their structures and functions. Metalexicographers claim that lexicography no longer is a sub-discipline of linguistics but an independent discipline. This is no uncontested or generally accepted claim and even many lexicographers often still see lexicographic theory as a mere application of linguistic theory. This paper argues that a significant contribution of e-lexicography is its confirmation of the status of lexicography as independent but also multidisciplinary field. This claim is substantiated by showing how the transition from printed to electronic dictionaries has influenced and continues to influence lexicographic theory.

2   A General Theory of Lexicography

When looking at e-dictionaries, we do not suggest an entirely new theory of lexicography directed exclusively at e-dictionaries, but rather opt for an adapted, extended theory that makes provision for both printed and e-dictionaries. The formulation of such a theory demands a thorough re-investigation of current lexicographic theories as well as the ways in which these theories need to

be adapted to have a general applicability. A modern general lexicographic theory should be anchored in three fundamental phases of the early development of lexicographic theory, see Gouws 2005, i.e. Zgusta's approach with the linguistic contents of dictionaries at the core, Wiegand's focus on dictionary structures and Bergenholtz and Tarp's emphasis on lexicographic functions. Relevant aspects from these phases should be supplemented and adapted to printed and e-lexicography to provide practical lexicographers with the necessary assistance when embarking on the compilation of a dictionary. In every theoretical approach and in all decisions the potential user and his/her needs and reference skills should remain the determining factor.

3   A Theoretical Basis for the Fast Growing e-Lexicography Practice

In printed and e-dictionaries for general languages, the linguistic contents will always remain important and the changes in lexicographic theory need to negotiate the best possible ways of selecting and presenting the linguistic data relevant to the intended target users of the dictionaries. In order to select the required linguistic data to be included as items in a dictionary, lexicographers need to be aware of the functions of the dictionary, and the presentation of the selected items demands a thorough utilisation of dictionary structures. Albeit that the linguistic contents and the functions of printed and e-dictionaries can be similar, the presentation differs and that demands innovative formulations of dictionary structures. For some traditional structures like the macro-, article and access structures a completely new approach is needed, see Gouws 2014a, 2014b. This paper focuses on some of these adaptations, compelled by a theory that is also directed at e-dictionaries. In one of the early discussions of lexicographic theory, Wiegand (1984:15) already stressed the importance of data collection, data processing and computer assistance. A new theory of lexicography is expanded by describing and taking cognisance of all innovative technological features enabled by the computer era, including data collection and processing and data representation and interface design, in order to respond to the different needs of the envisaged target users in an optimal way. This confirms that lexicographic theory needs more than a linguistic approach to ensure the best possible outcome.

4   Access on Demand to Different End-user Support Tools and

Additional Data

Major changes in the theory of lexicography expand the traditional notion of lexicography. A new theory is characterised by both an independent and an interdisciplinary nature. The theory looks beyond dictionaries as study field and includes issues such as access on demand to different end-user support tools and additional data. Where corpora previously assisted the lexicographer they can now also be accessed by the end-user. Any comprehensive theoretical model therefore has to make provision for, inter alia, the successful use of corpora for designing a dictionary (e.g. tools such as rulers and block systems), on-demand access by end-users to unprocessed and processed corpora (e.g. KWIC corpus examples, frequency trajectories and domain guidance interaction), user support tools (such as decision trees and direct user guidance tools), the linking of dictionaries to other reference sources and using dictionary databases in conjunction with other tools (such as natural language processing tools), such as writing assistants. The e-dictionary could, for example, link to a writing assistant in order to solve text production problems of users (Prinsloo, Prinsloo & Prinsloo 2015), and to provide specific information on demand. Lexicographic theory needs to provide for more than

mere traditional dictionary models. E-lexicography compels (meta-)lexicographers to negotiate various aspects, strategies and devices of additional user support, including all types of lexicographic tools, plugins, and linking with a variety of sources. This has direct implications for a variety of features of online dictionaries. The problem of data verification constitutes a vital component of a new approach to lexicographic theory. Additional information retrieved by means of a whole range of links is available, but users need to be guided or alerted to the trustworthiness of such information and a theory should describe these issues. It could firstly be done by formulating a set of guidelines to lexicographers regarding lemma selection and the balancing of alphabetical stretches in dictionaries, addressing the age-old problem of what to include and what to leave out. A lexicographic theory could secondly provide guidance with regard to the problems of over/under treating lemmas. Here the concept of lexicographic rulers (Prinsloo & De Schryver 2009) could be elaborated within a lexicographic theory. User support tools could be categorised as a taxonomy for a theory of e-dictionaries, as in Figure 1.
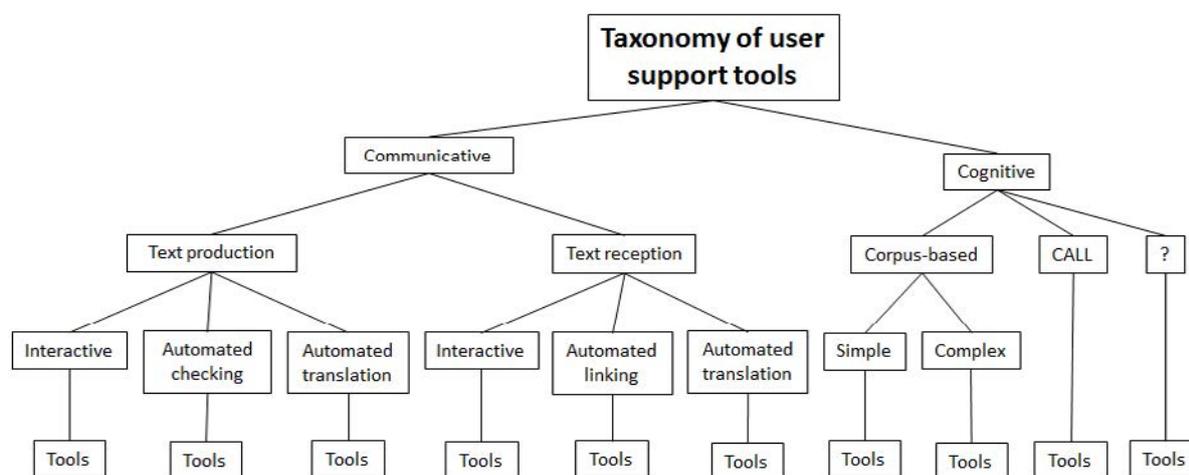
Figure 1: A taxonomy of user support tools.

The upper levels of the proposed taxonomy are based on the Function Theory of Lexicography, and the lower level on the nature of the interaction of the user with the tools. At the lowest level, the different tools are categorised, in terms of Prinsloo, Bothma & Heid 2015. Consider, for example, Figures 2-5 as examples from the categories text production, text reception and computer-aided language learning (CALL).

Figure 2: Text production with the *Sepedihelper* (Prinsloo, Prinsloo and Prinsloo 2015).

In Figure 2 a sentence in the situative mood in Sepedi is created through a step-by-step build-up combining the skills of the user with computational support in the complicated nominal and concordial system of Sepedi.
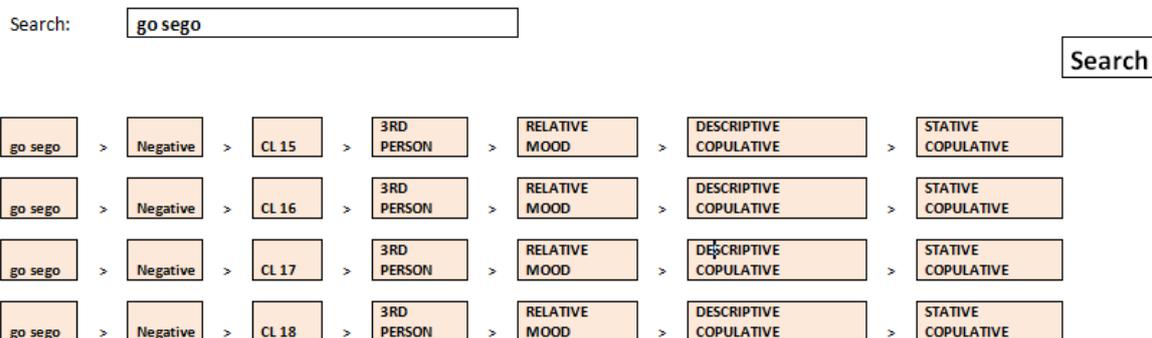
Figure 3: Text reception for Copulatives in Sepedi (Bothma, Heid and Prinsloo 2013).

Figure 3 gives a text reception breakdown of the different morpho-semantic possibilities of the copulative *go sego* in Sepedi.

Kollokasie	Relig.	Koerant
In diens	206	410
Jaar/jare diens	18	51
Beter/beste diens	0	65
Swak diens	0	55
Goeie diens	2	75
Aan diens	8	31
Onbaatsugtige diens	15	1
Besondere diens	180	0
Diakonale diens	15	0

Figure 4: Domain specific information (Prinsloo, Bothma and Heid 2013).

In Figure 4, guidance in terms of frequency of use in religious versus newspaper corpora is given for collocations of the Afrikaans word *diens*, 'service'.
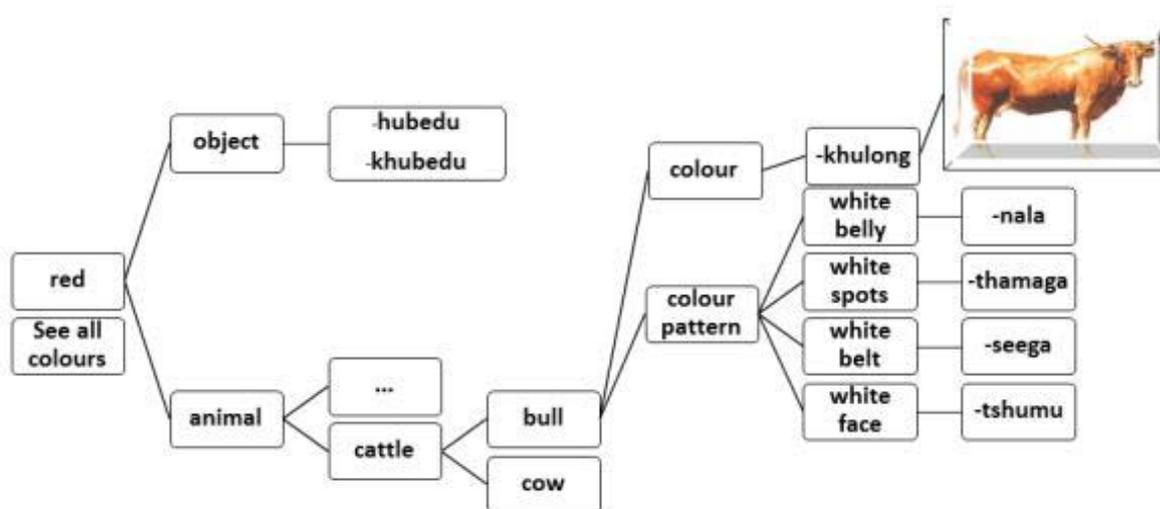
Figure 5 CALL: learning colour differences in Sepedi (Prinsloo, Taljard, Bothma and Heid 2013).

The decision tree in Figure 5 guides the user by means of a series of choices through the complicated colour system of Sepedi to the correct Sepedi word for a red cow.

5   A New Approach to the User Perspective

A theory that meets the demands of the practice of e-lexicography requires the development of a new approach to the user perspective. Insights from the field of information science facilitate the compilation of a much more sophisticated specification of user needs and the way in which users interact with the information (see Bergenholtz & Bothma 2011, Bothma 2011, Bergenholtz, Bothma & Gouws 2015).

Within the Function Theory of Lexicography (see Tarp 2008, Bothma & Tarp 2012, 2014, Fuertes-Olivera & Tarp 2014), the user needs in a given information need situation is paramount, as expressed in Bergenholtz & Gouws 2010: "For the user the type of information source is not important. Important is that he/she retrieves the exact required information as quickly as possible". Sophisticated searching and filtering technologies, often combined with profiling, are commonly used in information science to enable exactly this (Bothma 2011). Through advanced search interfaces and the specification of display options, the user can limit the amount of information (s)he is confronted with after a search, and therefore ensure that there is no information overload.

Two examples will suffice:

a)   A user may need to find examples of the usage and senses / meanings of a specific word in poetry published in the 18th and 19th centuries and (s)he would like to see only the specific examples from the specified corpus. This would typically require a once-off profile, in which the user searches for the word, but indicates that the word should occur in examples from the specified dates and genres; the user could furthermore specify that only the senses / meanings in which the specified word occurs in poetry should be displayed on screen.

b)   A translator doing highly technical translation from Spanish into Danish may only want

technical equivalents (not general terms used colloquially), combined with explanations at a high level of complexity and detail, in addition to examples from technical literature. This would enable the user to access exactly the correct data in terms of his/her information needs while translating, and save a huge amount of time not wading through non-technical / popular translation equivalents and examples.

Compiling such dictionaries would require modifications to the database structure and both the search and filtering options of e-dictionaries, as well as the display options:

- The database should be much more granular and contain much more metadata. In the above examples it would, for example, require the following:
  - In the first example, each corpus example would need a metadata field "genre", as well as "date".
  - In the second example of translation dictionaries, translation equivalents would need a field indicating whether the specific translation equivalent is colloquial or technical; the same applies to the corpus examples. In addition, explanations would need a metadata field specifying whether the explanation is aimed at a lay person, a semi-expert or an expert.
- These metadata fields should be accessible to set as both search and filtering parameters:
  - In the first example, the user should have the option to specify "genre" and "date" as filtering mechanisms.
  - In the second example, the user should have the option to specify "technical" and "expert level of complexity" as filtering mechanisms.
- Based on the user's information needs, only the required data should be displayed on screen. The user should therefore have the option of specifying which data are to be displayed:
  - In the first example, the user should be able to specify that only the meanings / senses and concomitant corpus examples from the specified date range be displayed on screen, and not the full dictionary article with all its corpus examples.
  - The same applies to the second example, in which case all non-technical translation equivalents and corpus examples, as well as non-expert level explanations, should be filtered out.

An advanced search interface, combined with specifications on what should be displayed on screen, can easily facilitate presenting exactly the right information to the user. Such an advanced search with the specification of what should be displayed is in fact a once-off specification of an active user profile.

A profile can, however, be set as a persistent active profile as well (Bothma 2011). In an active profile, the user sets the parameters for his/her information needs, either as a once-off option (as indicated above), or as a persistent option, for example by checking radio buttons that specify the specific information need, where the system should search and what data should be displayed, and this profile is then stored in the system and used each time the user executes a search. Such a profile enables rapid access on demand to the selection of microstructural items dependent on the user's specific information needs. Obviously, it must always be possible to change one's user profile in terms of search and display parameters, depending on changing information needs over time (Bothma & Bergenholtz 2013). (Profiling can also be a passive user profile. In the case of a passive profile, the system tracks the user's searching and navigation behaviour, and on the basis of this, builds up a profile of the user's preferences (Bothma 2011).)

It is therefore possible to present to the user the information (s)he needs in a specific situation at the required level of detail and complexity (Bergenholtz & Bothma 2011), for example, detailed,

complex information targeted at the expert, or a brief, simplified version targeted at the lay person. This also requires additional metadata fields, as indicated above.

By means of the tools and technologies described above, users are able to access information and support on demand – additional support is available when needed, but users are not automatically being subjected to an overload of information.

No such e-dictionaries that the authors are aware of currently exist. Such dictionaries will, however, comply with Tarp's notion of the "'individualization' of the lexical product, adapting to the concrete needs of the concrete user" (Tarp 2009a), within the Function Theory of Lexicography, thereby satisfying the information needs "that only show up in one out of a hundred or one out of a thousand consultations" (Tarp 2009b).

With such a complexity of support tools and search and display options, user acceptance of the electronic interface becomes extremely important. Usability studies therefore need to be conducted to ascertain to what extent users make use of such complex search, filtering and display options, and how to optimise the display options on the screen. Aspects of human-computer interaction, usability engineering / usability studies will therefore also make important contributions towards the development of such sophisticated e-dictionaries.

To be able to compile such e-dictionaries a theory is needed that negotiates a vast range of new components that were not relevant in either linguistic theory or in lexicographic theory directed exclusively at printed dictionaries. Such an approach confirms the independent and multidisciplinary nature of lexicographic theory.

6 Conclusion

The transition from printed to e-lexicography has ascertained the interdisciplinary nature of lexicography as an independent and multidisciplinary field. Different disciplines, e.g. human language technology, linguistics and information science, become increasingly important role-players in the development of a lexicographic theory directed at the practice of dictionary compilation. A redefined theory of lexicography needs to reflect these contributions but should also display a comprehensive nature providing a sound theoretical basis to the lexicographic practice that responds to the needs of diverse and wide-ranging potential target user groups.

7 References

Bergenholtz, H. & Bothma, T.J.D. 2011. Needs-adapted data presentation in e-information tools. *Lexikos* 21: 53-77.

Bergenholtz, H., Bothma, T.J.D. & Gouws, R.H. 2015. Phases and steps in the access to data in information tools. *Lexikos* 25: 1-30.

Bergenholtz, H. & Gouws, R.H. 2010. A New Perspective on the Access Process. *Hermes. Journal of Language and Business Communication* 44: 103-127.

Bothma, T.J.D. 2011. Filtering and adapting data and information in the online environment in response to user needs. In Fuertes-Olivera, PA and Bergenholtz, H. (Eds.) 2011. *e- Lexicography: The Internet, Digital Initiatives and Lexicography.* London & New York: Continuum, 71-102.

Bothma, T.J.D. & Bergenholtz, H. 2013. "Information needs changing over time": a critical discussion. *SAJLIS* 79(1): 22-34.

Bothma, T.J.D. & Tarp, S. 2012. Lexicography and the relevance criterion. *Lexikos* 22: 86-108.

Bothma, T.J.D. & Tarp, S. 2014. Why relevance theory is relevant for lexicographers. *Lexicographica* 30: 350-378.

Bothma, T.J.D., Heid, U. & Prinsloo, D.J. (2013). Implementing decision support for text production in e-dictionaries. *eLex 2013*, Tallinn, 17-19 Oct. 2013.

Fuertes-Olivera, P.A. & Tarp, S. 2014. Theory and practice of specialised online dictionaries: Lexicography versus terminography. Berlin / New York: De Gruyter.

Gouws, R.H. 2005. Meilensteine auf dem historischen Weg der Metalexikographie. *Lexicographica* 21: 158-178.

Gouws, R.H. 2014a. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.

Gouws, R.H. 2014b. Makrostruktuuraanpassings vanaf gedrukte na e-woordeboeke. *Tydskrif vir Geesteswetenskappe* 54 (3): 481-504.

Prinsloo, D.J. & de Schryver, G-M. 2009. Designing a Measurement Instrument for the Relative Length of Alphabetical Stretches in Dictionaries, with special reference to Afrikaans and English. In Braasch, A. & Povlsen C. (eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002*, Copenhagen, Denmark, August 13-17, 2002: 483-494. Copenhagen: Center for Sprogteknologi, Københavns Universitet.

Prinsloo, D.J., Bothma, T.J.D. & Heid, U. (2013). Linking e-dictionaries to advanced processed corpus data. *eLex 2013*, Tallinn, 17-19 Oct. 2013.

Prinsloo, D.J., Bothma, T.J.D. & Heid, U. 2015. User guidance devices for e-lexicography. *SeLA Conference.* University of Hildesheim, 7-8 October 2015.

Prinsloo, D.J., Prinsloo, Daniel and Prinsloo, Jacobus. A writing tool for Sepedi. *E-Lex 2015.* Herstmonceux Castle. United Kingdom. 11-13 August 2015.

Prinsloo, D.J., Taljard, E., Bothma, T.J.D., & Heid, U. (2013). Interactive decision trees for teaching and learning of Northern Sotho colour terms. *HLT4LL 2013 Workshop.* Stellenbosch, 11-12 Nov. 2013.

Tarp, S. 2008. Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography. Tubingen: Niemeyer.

Tarp, S. 2009a. Reflections on Data Access in Lexicographic Works, in S. Nielsen & S. Tarp (eds). 2009. *Lexicography in the 21st Century. In honour of Henning Bergenholtz.* Amsterdam / Philadelphia: John Benjamins, 43–62.

Tarp, S. 2009b. Reflections on Lexicographical User Research. *Lexikos* 19: 275-296.

Wiegand, H.E. 1984. On the structure and contents of a general theory of lexicography. In Hartmann, R.R.K. (ed.) 1984. *LEXeter '83.* Tübingen: Max Niemeyer: 13-30

Acknowledgements