

# The microstructure of Online Linguistics Dictionaries: obligatory and facultative elements

Carolina Flinz

University of Pisa

Via S. Maria 36, 56100 Pisa

E-mail: c.flinz@ec.unipi.it

## Abstract

The planning of a dictionary should consider both theoretical and empiric aspects, either for its macro- and microstructure: this is true also for Online Specialized Dictionaries of Linguistics. In particular the microstructure should be standardized and structured so as to fit with the primary and secondary functions of a dictionary. Unfortunately, empirical studies that investigate Online Specialized Dictionaries of Linguistics are rare, making it unclear which microstructural elements are obligatory and which are facultative. This article will present and comment upon the results of an investigation into a corpus of Online Specialized Dictionaries of Linguistics, focusing attention on these aspects and also the most important theoretical issues. An example taken from DIL, a German-Italian Online Dictionary of Linguistics, will end the article.

**Keywords:** online dictionaries; LSP dictionaries; online dictionaries of linguistics; microstructure bilingual dictionaries of linguistics

## 1. Introduction

Language Specific Dictionaries are both central to research and core components of basic literature. However, they have only recently become objects of scientific attention: prior to the end of the twentieth century, few lexicographers focused their attention on specialized register<sup>1</sup> and its principal characteristics<sup>2</sup> (cf. Schaefer & Bergenholtz, 1994).

Dictionaries of linguistics are also LSP dictionaries. As such, they have only been investigated since the beginning of the twenty-first century<sup>3</sup> - despite their position as important and relevant tools for the scientific community (in both their printed and online forms). Still today, empirical analyses remain almost non-existent, meaning that no guidelines are present for future LSP lexicographers with an interest in this field.

A particular need also exists for bilingual dictionaries of linguistics in Italy (and particularly German/Italian versions). This is due to the 1999 Italian program of university reforms which separated language studies from literature studies - in turn giving a new role to foreign language linguistics.

This article aims to present the results of an analysis carried out on the microstructure of existing online dictionaries of linguistics. The analysis was originally carried out with the intention of investigating these dictionaries' principal microstructural characteristics<sup>4</sup> - in the hope of providing a set of guidelines for future LSP lexicographers.

Online dictionaries of linguistics are works in their own right, and so should not be seen as simply web-based versions of printed works. Instead, they have their own rights (Barz, 2005), even if some terminology overlaps with printed versions; online dictionaries can be terminology banks, language learning environments (Störrer, 1998), and working and discussion platforms (Abel, 2006). The limits between dictionary, archive, grammar and databank are not strict, and the extensive use of the terms "glossary" and "lexicon" is very common (Flinz, 2010:71).

The microstructure of electronic dictionaries of linguistics - and especially online versions - has only been investigated since the beginning of the twenty-first century. A brief overview of the current state of art will be undertaken in Section 2.

The corpus of the analyses will then be presented in Section 3, and the commented results and some specific examples will end the paper in Section 4. Obligatory and facultative elements will also be focused upon.

## 2. State of the art: microstructure

The microstructure of dictionaries has long been a point of discussion in lexicography. Many authors have tried

<sup>1</sup> „Was ist eigentlich Fachlexikographie?“ (Wiegand, 1988) is considered to be one of the first articles written about this theme. Still in the 80s many lexicographers considered this field “unorganized” and used terms like “vegetating state” (Kucera, 1984), “Wildwuchsgebiet” (Wiegand, 1988) even if there was a “shifting towards the specialized register” (Pilegaard, 1994). The most important publications are from the 90s (Dressler & Schaefer, 1994; Schaefer & Bergenholtz, 1994; Bergenholtz & Tarp, 1995; Hoffmann et al., 1998; Hoffmann et al. 1999 etc.)

<sup>2</sup> Even if the total amount of published LSP dictionaries is very high (Cf. Dressler, 1994), the ones published for a single discipline are very few (excepts some fields like medicine etc.).

<sup>3</sup> Cf. Adamzik, 2001; Lorenzi, 2002; Kreuder, 2003; Flinz, 2010.

<sup>4</sup> The results have been used to plan the microstructure of a German-Italian online dictionary linguistics - a project at the University of Pisa. The results of the analyses concerning the macrostructure were published in elex 2009.

to give the topic a precise definition:

- 1) Rey-Debove defines it as the "ensemble des informations ordonnées de chaque article [...] à la suite de l'entrée" (Rey-Debove 1971:13);
- 2) Hausmann % Wiegand (1989:328) consider it as "the structure of information within the article".

Dictionary microstructures are made up of several different elements that offer information regarding entries' semantics and form (Wiegand, 1996). This composition permits a range of possibilities (cf. Zöfgen, 1994:108):

"Lemma:

- a) Aussprache, grammatische Angaben (Artikel, Genus, Pluralbildung, Deklination, Konjugation ...), Markierung;
- b) (Polysemieangaben: 1°, 2°,...); Bedeutungsparaphrase, Syntagmatik, Valenzangaben, Kollokationen, Beispiele, Paradigmatik, Synonyme, Antonyme, Begriffsfelder, (Homonymie,...);
- c) Phaseologische Angaben / Idiomatik."<sup>5</sup>

Of these possibilities, Wiegand distinguishes the three most common variations:

- 1) „integrierte Mikrostruktur“, which gives all of the information in single polisemic entries;
- 2) „gemisch-integrierte Mikrostruktur“, which contains each of the polisemic variations in each entry, with each meaning being followed by its own information;
- 3) „gemisch-semiintegrierte Mikrostruktur“ (Wiegand, 1996), which has both integrated and unintegrated parts, and is generally used for larger articles.

The number and type of information that should follow each entry is determined by the dictionary's type and function. Theoretical studies include both formal and encyclopedic information in LSP dictionaries, as: 1) this information is usually missing in language dictionaries that focus on general language; and 2) this information can help users to better understand lexical lacunae and partial correspondence of meaning.

Standardizing the microstructure is also crucial, as doing so:

- 1) ensures ease of comprehension;
- 2) reduced the time required to find information;
- 3) creates a homogenous style throughout the dictionary;
- 4) limits individual entries' wording;
- 5) simplifies the dictionary's readability.<sup>6</sup>

Taken together, these points improve user interaction,

<sup>5</sup> In English: „Entry: a) pronunciation, grammatical informations (article, genre, plural, declension, coniugation...), marks; b) Polysemy (1°, 2°,...); paraphrase of meaning, syntagmatic informations, valency informations, collocations, examples, paradigmatic informations, synonyms, antonyms, conceptual fields (homonymity); c) phraseological informations /idiomatic”.

<sup>6</sup> Cf. Wiegand, 1989; Adamzik, 2001.

causing the “Lesewörterbuch” to be positively judged (Kühn, 1998). Further efforts in this regard include:

- 1) limited use of links, as too many can cause readers to feel lost;
- 2) careful use of LSP words, which would not be understandable to the average reader;
- 3) limited use of abbreviations, saving the reader from regularly having to check up on meanings.

### 3. Corpus

The corpus is made up of 27 linguistics dictionaries that were found with the help of search engines such as Google, AltaVista and Lycos.

These dictionaries were then categorized in the following manner:

- 1) monolingual (11 in English, 6 in German, 0 in Italian);
- 2) bilingual (6 with English as L1 and German, French, Spanish as L2; 1 with German as L1 and English; 1 with Russian as L1 and German as L2; 0 with Italian)
- 3) plurilingual (2 with Italian as L1).<sup>7</sup>

The dictionaries' microstructure was analyzed according to:

- 1) article header: typographic relevance of the entry; number and type of information (phonetic, grammar, domain);
- 2) equivalent or equivalents: languages; direction; presence of grammatical information about the equivalent;
- 3) definition: use of paraphrasing or citations; presence of links; presence of specific language; presence of abbreviations;
- 4) syntagmatic elements: syntax information; collocations; examples;
- 5) paradigmatic elements: synonyms; indication of words belonging to the same semantic field;
- 6) bibliographical information.

### 4. Results

The dictionaries were analyzed according to the above-cited categories. The following general considerations were also taken into account:

- 1) three of the dictionaries were only very simple glossaries – giving equivalents in the foreign language, but no indication of synonymy and/or related terms. Similarly, they fail to typographically mark and define each entry, don't include bibliographical information, and list every polisemic meaning on its own. They are also organized in a very simple manner, without the typical elements found in online dictionaries (search engine, links etc.);
- 2) almost all dictionaries had English as L1: only in one case was German used (in a bilingual dictionary) and

<sup>7</sup> In this abstract, only the results pertaining to bilingual dictionaries will be presented.

Italian (in a plurilingual dictionary);

3) there was great variation between L2 in the bilingual dictionaries: German, Spain, French, Chinese and Russian were all found. It therefore wasn't possible to compare two dictionaries with the same language couple, meaning that only one dictionary for each language pair has been included;

4) 50% included one entry for each polisemic variation of each term, and 50% had one entry with the various meanings of the lemma (usually marked with the help of letters or numbers).

#### 4.1. Article Header

83% of the analyzed dictionaries had a typographically marked entry in black and bold, although one also included entries in red. In another dictionary, entries are marked with a larger font-size.

There were generally two pieces of information contained in each article's header – most frequently word class and informations about synonymy.

More detailed results are given in the following diagram:

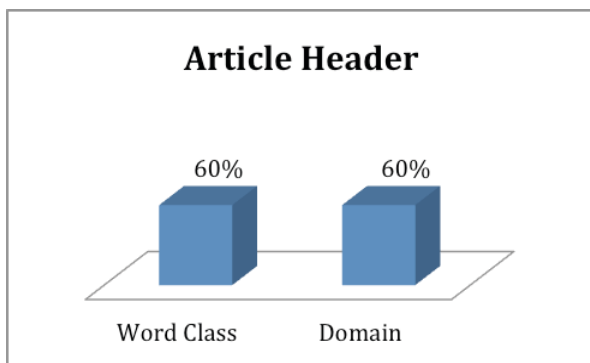


Figure 1: Results from the analyses of the article header

60% of the analyzed dictionaries give information about word class (noun, verb, adjective etc.) and specific domain (lexicography, applied linguistics etc.).

#### 4.2. Equivalentents

Equivalentents were the most typical elements found in bilingual dictionaries. The most common source language was English, with German and Italian being found as the source language only once.

Bidirectional dictionaries are generally considered rare in theoretical papers, but the results of this analysis do not confirm this fact: 66% of the analyzed dictionaries were bidirectional, even if in some cases this was not total bidirectionality. Both products from the Summer Institute of Linguistics are good references for this type of dictionary: the user of the "French/English Glossary of Linguistics Terms", for example, can chose on the first page between entries that are French, English or bidirectional.

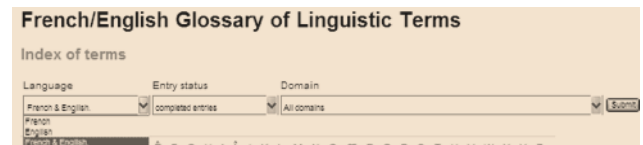


Figure 2: Screenshot from French/English Glossary of Linguistic Terms (www.sil.org)

Lexicographers also usually consider the inclusion of grammatical information to be common, but this was only found rarely in the dictionaries analyzed (8%).

#### 4.3. Definition

Definition is an important and relevant element in encyclopedic dictionaries, and also in LSP dictionaries (Rossenbeck 1987: 278f; Duvå–Laursen 1994:247f; Schaefer & Bergenholtz, 1994:141). The analyses showed that 83% of the dictionaries were equipped with definitions.

The definition is considered "the epicenter of the microstructure" (Schaefer & Bergenholtz, 1994:225). As such, the primary consideration should be that "the preciseness, the scope and currency of the explanation" should be coupled with "an up-to-date comment on their specific usage" (Schaefer & Bergenholtz, 1994:219). Lexicographers should always have in mind who the user is and what his needs are Kühn (1998), Wiegand (2002) und Zöfgen (1994).

Different possibilities exist for drafting definitions:

- 1) use of the Aristotle principle of *genus proximum* and *differentia specifica*;
- 2) information of the intention and the extension of the term;
- 2) indication about the most important prototypical semantic aspects;
- 3) evidence about the principle related concepts ("frame concepts").

Similarly, various techniques can be used in the writing process:

- 1) paraphrasing;
- 2) including synonyms;
- 3) contrastive analyses;
- 4) including citations.

The results of the analyses showed that 60% of the dictionaries used paraphrasing in their explanation, and 40% included citations. The Spanish-German dictionary from Hispanoteca and the DLM Project gave citations in both relevant languages or more.

Each definition should present different perspectives on terms, evidenced by the use of paragraphs. (Adamzik, 2001:220). However, the use of paragraphs in the analyzed dictionaries is not common: the only exception is the Spanish-German dictionary from Hispanoteca. A key concept in lexicography is that words should be

defined using words simpler than themselves (Götz, 1984: 50) – but 100% of the analyzed dictionaries used LSP words in their definitions, even if the proportion was limited to a few words (3-5) per entry. Abbreviations are rare (despite being a typical element of printed linguistics dictionaries).

Links and references to other related terms are very common, being present in all dictionaries with the exception of the Spanish-German dictionary from Hispanoteca. This exception seems to be a transposition of a written dictionary, having hyperlinks neither in the text nor at its end. Instead, it uses the method commonly found in printed dictionaries:

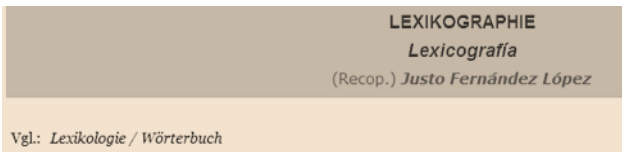


Figure 3: Screenshot from Lexikon der Linguistik / Diccionario de Lingüística (www.hispanoteca.eu)

#### 4.4. Sintagmatic Information

Sintagmatic information is very important in some types of dictionary (learner’s dictionaries, language dictionaries and LSP dictionaries), because it shows how an entry can be used and what types of words are accepted (Zöfgen, 1994:147). Theories distinguish between:

- 1) syntactical information, which focuses on the correct use of a term;
- 2) collocations: the union of two or more words in a sentence. These are also welcomed in bilingual dictionaries;
- 3) examples, which usually show what type of constructions can be built with the lemma. Examples are also used to show concepts or give particular information.

The results of these points are demonstrated in the following diagram:

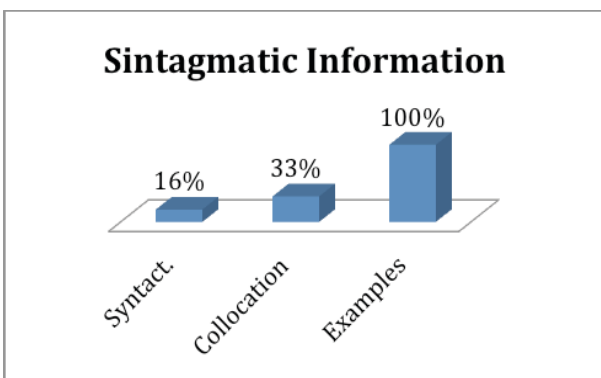


Figure 4: Results from the analyses of the sintagmatic information

Syntactical information is rare, only being used in one dictionary, and even then including little detail. Similarly, information about collocations is not used much in this type of dictionary, with the only exception being the Linguistics Glossary. At the end of each entry here is found the indication: (Concordances for ...):

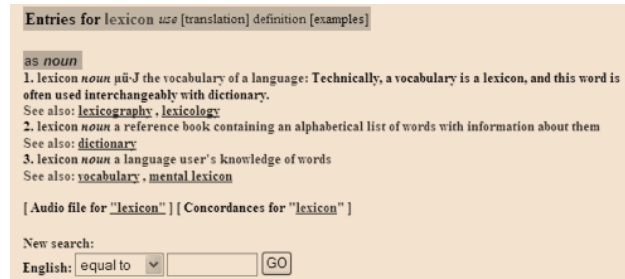


Figure 5: Screenshot from Linguistics Glossary (www.edict.biz/lexiconindex/linguistics)

On opening the link to Concordances for Lexicon, more search options are presented:

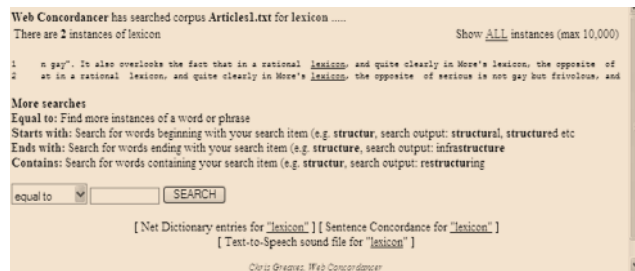


Figure 6: Screenshot from Linguistics Glossary (www.edict.biz/lexiconindex/linguistics)

An option also exists for searching sentence concordance:

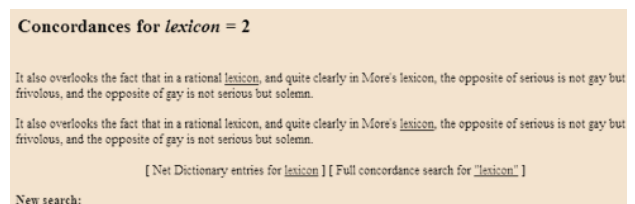


Figure 7: Screenshot from Linguistics Glossary (www.edict.biz/lexiconindex/linguistics)

Examples are present in 100% of the analyzed dictionaries. They can be:

- 1) examples of use;
- 2) examples of concepts (e.g. pronunciation in the definition of particular types of vowels: the “Glossary Spanish-English” has audio files that let the user better understand the audio characteristics of specific vowels).

#### 4.5 Paradigmatic Information

Paradigmatic information about entries is welcome in dictionaries (Kühn, 1998). It is separated into: 1)

synonyms; 2) antonyms; 3) terms belonging to the same semantic area.

The analyses concluded the following results:

- 1) synonyms are present in all dictionaries;
- 2) antonyms are used only in one dictionary;
- 3) related terms are typical of these dictionaries: 100% introduce related terms in their entries, which are usually included at either the beginning or end of articles and are signaled with “see also”, “cfr” etc.

#### 4.6. Bibliographical Information

Bibliographical information is a typical element of German dictionaries of linguistics, where each entry has at its end a complete indication of the source. These are absent in Italian dictionaries, which instead include them at the dictionaries’ ends.

The analyses of online dictionaries showed that only 66% put this type of information at the end of the articles. The dictionaries from the Summer Institute of Linguistics put them in abbreviated form, with a link allowing users to open the full information.

Theoretical studies consider the following as key components of linguistic dictionaries: abbreviations, synonyms, information about grammatical aspects and used contexts, antonyms, etymology and examples. The

Example:

empirical analyses confirmed these considerations only partially, with bilingual linguistics dictionaries being shown to have:

- 1) lexicographically-marked entries;
- 2) indication about word class and domain in the article header, but further grammatical information rarely attested even if strongly called for by many researchers;
- 3) examples of concepts;
- 4) synonyms and related terms.

The online dictionaries also have a good structured definition: they alternate the use of paraphrasing and citations. Differing perspectives on entries are shown with the help of paragraphs. LSP words, links and abbreviations are also used in the right way, without disturbing the user in his purposes.

These analyses were a great help in the planning of a bilingual German-Italian linguistics dictionary at the University of Pisa (naturally after considering the potential user, his needs, his probable situation of use, and also the type and function of the dictionary).

The microstructure of DIL, a German-Italian online Specialized Dictionary of Linguistics, has tried to follow the above criteria:

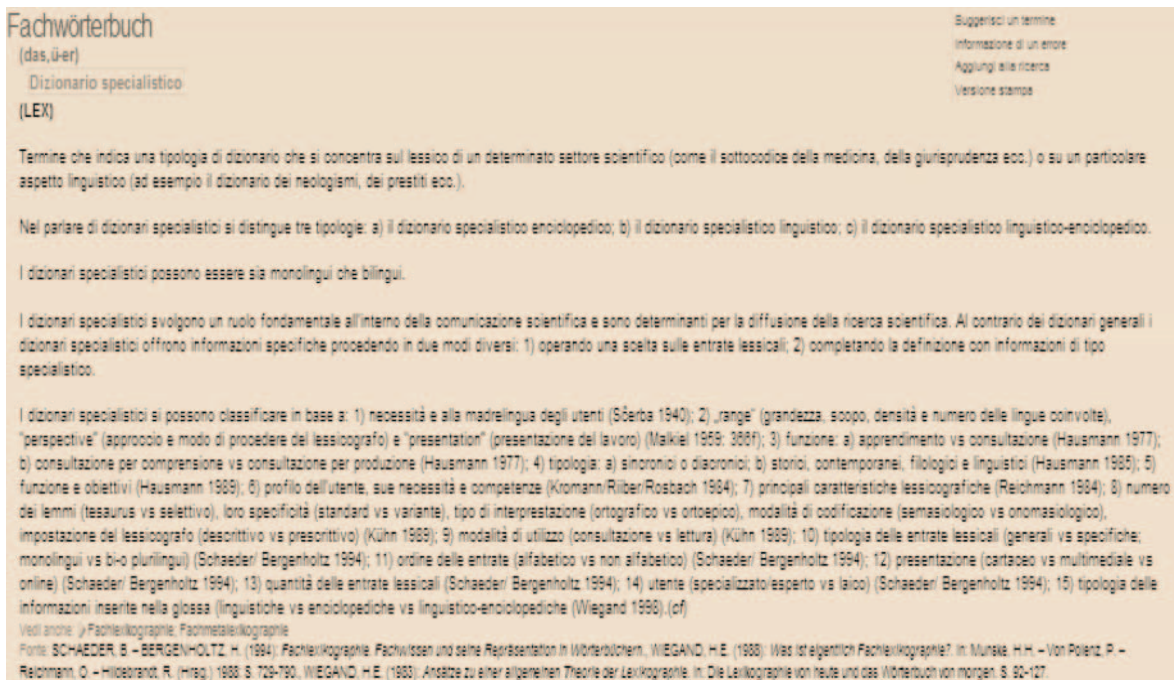


Figure 8: Screenshot from DIL / Dizionario Tedesco-Italiano di terminologia linguistica) (www.humnet.unipi.it/dott\_linggensac/glossword)

## 5. References

- Abel, A. (2006). Elektronische Wörterbücher: Neue Wege und Tendenzen. In F. San Vicente (ed.) *Akten der Tagung „Lessicografia bilingue e Traduzione: metodi, strumenti e approcci attuali (Forlì, 17.-18.11.2005)“*, Open Access Publications: Polimetrica Publisher, pp. 35-56.
- Adamzik, K. (2001). *Kontrastive Textologie*, Tübingen: Stauffenburg.
- Barz, I., Bergenholtz, H. & Korhonen, J. (2005). *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*. Frankfurt a.M.: Lang.
- Bergenholtz, H., Schaefer, B. (1994). *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Narr.
- Bergenholtz, H., Tarp, S. (1995). *Manual of LSP lexicography. Preparation of LSP dictionaries-problems and suggested solutions*, Amsterdam - Philadelphia: John Benjamins.
- Dressler, S. (1994). Wörterbücher der Medizin. Eine Bibliographie. In S. Dressler, B. Schaefer (eds.) *Wörterbücher der Medizin. Beiträge zur Fachlexikographie*. Tübingen: Niemeyer, pp. 171-279.
- Dressler, S., Schaefer, B. (1994). *Wörterbücher der Medizin. Beiträge zur Fachlexikographie*. Tübingen: Niemeyer. (Lexikographica Series Maior 55).
- Duvå, G., Laursen, A.L. (1994). Translation and LSP Lexicography: A User Survey. In H. Bergenholtz, B. Schaefer (eds.) *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Narr, pp. 247-267.
- Flinz, C. (2010). DIL- an online bilingual specialized dictionary of linguistics (German-Italian). In S. Granger, M. Paquot (eds.) *eLexicography in the 21st century: New challenges, new applications*. Louvain: UCL, pp. 67-77.
- Götz, D., Herbst, T. (1984) (eds.). *Theoretische und praktische Probleme der Lexikographie. Professor Dr. Günther Haensch zum 60. Geburtstag I. Augsburger Kolloquium*, München: Hueber.
- Hausmann, F., Reichmann, O., Wiegand, H.E. & Zgusta, L. (eds.) (1989). *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Handbücher zur Sprach- und Kommunikationswissenschaft (HSK 5.1.) Berlin, New York: De Gruyter.
- Hausmann, F., Reichmann, O., Wiegand, H.E. & Zgusta, L. (eds.) (1990). *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Handbücher zur Sprach- und Kommunikationswissenschaft (HSK 5.2.) Berlin, New York: De Gruyter.
- Kreuder, H.D. (2003). *Metasprachliche Lexikographie. Untersuchungen zur Kodifizierung der linguistischen Terminologie*, Tübingen: Niemeyer.
- Kucera, A. (1984). Aus der Werkstatt der praktischen Verlagslexikographie. Übersetzungswörterbücher der Fachsprachen. In: *Mitteilungen für Dolmetscher und Übersetzer* 1/30. pp. 3-6.
- Kühn, P. (1998). *Langenscheidts Großwörterbuch Deutsch als Fremdsprache und die deutschen Wörterbücher*. In H.E. Wiegand (ed.) *Perspektiven der pädagogischen Lexikographie des Deutschen. Untersuchungen anhand von "Langenscheidts Großwörterbuch Deutsch als Fremdsprache"*, Tübingen: Niemeyer, pp. 34-60.
- Lorenzi, F. (a cura di) (2002). *DLM - Dizionario generale plurilingue del Lessico Metalinguistico*. Roma: Il Calamo.
- Pilegaard, M. (1994). Bilingual LSP Dictionaries. User benefit correlates with elaborateness of „explanation“. In H. Bergenholtz, B. Schaefer (eds.) *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Narr, pp. 211-228.
- Rey-Debove, J. (1971). *Étude linguistique et sémiotique des dictionnaires français contemporains*. Paris: The Hague.
- Rossenbeck, K. (1987). Zur Gestaltung zweisprachiger Fachwörterbücher. In A.M. Cornu, J. Vanparijs, M. Delahaye & L. Baten (eds.) *Beads or Bracelet? How do we approach LSP. Selected Papers from the Fifth European Symposium on LSP* (1985). Oxford: Oxford University Press, pp. 274-283.
- Storrer, A., Harriehausen, B. (1998). *Hypermedia für Lexikon und Grammatik*. Tübingen: Narr.
- Welker, H.A. (2003). *Zweisprachige Lexikographie: Vorschläge für deutsch-portugiesische Verbwörterbücher*. München: Utz.
- Wiegand, H.E. (1988). Was ist eigentlich Fachlexikographie?. In H.H. Munske, P. Von Polenz, O. Reichmann & R. Hildebrandt (eds.) *Deutscher Wortschatz. Lexikologische Studien*. Berlin & New York: De Gruyter, pp. 729-790.
- Wiegand, H.E. (1989). Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. In F.J. Hausmann, O. Reichmann, H.E. Wiegand & L. Zgusta (eds.), *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Handbücher zur Sprach- und Kommunikationswissenschaft HSK 5.1, pp. 409-462.
- Wiegand, H.E. (1996). *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger Lexikographie-Kolloquium*. Tübingen: Niemeyer.
- Wiegand, H.E. (ed.) (2002). *Perspektiven der pädagogischen Lexikographie des Deutschen. Untersuchungen anhand des „De Gruyter Wörterbuch Deutsch als Fremdsprache“*. Tübingen: Niemeyer.
- Zöfgen, E. (1994). *Lernerwörterbuch in Theorie und Praxis. Ein Beitrag zur Metalexikographie mit besonderer Berücksichtigung des Französischen*. Tübingen: Niemeyer.