
Towards a Framework for the Description and Evaluation of Dictionary Evaluation Criteria*

Piet Swanepoel, *Department of Afrikaans and Theory of Literature,
University of South Africa, Pretoria, Republic of South Africa*
(swaneph@unisa.ac.za)

Abstract: The reviewing of dictionaries is a time-honoured praxis, but reviewers have often been criticized for only describing the design features of the dictionaries under review and not evaluating these features, for narrowing their reviews down to one or two subjectively chosen design features, and for not explicating their evaluation criteria or indicating their validity.

Improving the quality of dictionary reviews is a task that has been delegated to dictionary criticism as a component of metalexigraphy. In this article, it is proposed that a first step in this direction would be to devise for each dictionary type a set of comprehensive, objective, valid, generally accepted, and operationalizable evaluation criteria. To this end, a general framework for the description and evaluation of dictionary evaluation criteria is developed in this article, using parameters from research on dictionary criticism and the usability of websites.

Keywords: DICTIONARY CRITICISM, DICTIONARY EVALUATION, DICTIONARY EVALUATION CRITERIA, DESCRIPTION OF DICTIONARY EVALUATION CRITERIA, EVALUATION OF DICTIONARY EVALUATION CRITERIA

Opsomming: 'n Voorstel vir 'n raamwerk vir die beskrywing en evaluering van woordeboekevalueringskriteria. Die resenseer van woordeboeke is 'n gerespekteerde praktyk, maar resensente is al dikwels daarvan beskuldig dat hulle net 'n beskrywing gee van die ontwerpkenmerke van woordeboeke onder bespreking en nie hierdie kenmerke evalueer nie, dat hulle hul resensies beperk tot een of twee subjektief gekose ontwerpkenmerke en dat hulle nie hulle evalueringskriteria duidelik uiteensit of hulle geldigheid aandui nie.

Die verbetering van die gehalte van woordeboekresensies is 'n taak wat opgedra is aan die woordeboekkritiek as 'n onderdeel van die metaleksikografie. In hierdie artikel word voorgestel dat 'n eerste stap in hierdie rigting sou wees om vir elke woordeboektype 'n stel omvattende, objektiewe, geldige, algemeen aanvaarde en operasionaliseerbare evalueringskriteria te ontwikkel. Vir hierdie doel word daar in hierdie artikel 'n algemene raamwerk vir die beskrywing en beoordeling van woordeboekevalueringskriteria ontwikkel deur gebruikmaking van parameters van navorsing oor woordeboekkritiek en die bruikbaarheid van webwerwe.

* This article is an extended version of a paper presented at the Thirteenth International Conference of the African Association for Lexicography, organized by the Bureau of the Woordeboek van die Afrikaanse Taal, Stellenbosch, Republic of South Africa, 30 June–3 July 2008. I would like to thank Rufus Gouws, Howard Jackson and Reinhard Hartmann for their comments on the first draft of this article.

Sleutelwoorde: WOORDEBOEKKRITIEK, WOORDEBOEKEVALUERING, WOORDEBOEK-EVALUERINGSKRITERIA, BESKRYWING VAN WOORDEBOEKEVALUERINGSKRITERIA, EVALUERING VAN WOORDEBOEKEVALUERINGSKRITERIA

1. Introduction

Legitime und gute Wörterbuchkritik muss den Werken voraus sein, die sie kritisiert. Sie muss geradezu die Umrisse neuer Werke erschaffen. Solche Wörterbuchkritik ist produktiv und selbst ein Teil der kritischen Wörterbuchforschung. (Wiegand 1994: 3)

Ideally seen, dictionary criticism as praxis is the evaluation, negative or positive, of the design features of a dictionary/dictionaries on the basis of one or more lexicographically relevant evaluation criteria (Ripfel 1989: 49 ff.). Although dictionary criticism/evaluation is a time-honoured practice (cf. Hartmann 1996: 241), it has come under some heavy criticism. Dictionary reviews have been described in derogatory terms as 'primitive' (Béjoint 2000: 113), as often being nothing more than 'incidental sniping' at unmotivated aspects of the design of dictionaries (Osselton 1989: 229), and as being no more than 'repetitions of the publisher's own publicity on the dictionary cover' (Bergenholtz and Tarp 1995: 232).

Even more substantial reviews have been criticized. It has been claimed

- that dictionary reviews often only describe the design features of the dictionary/dictionaries under review and seldom evaluate them (cf. Chan and Taylor 2001, and Osselton 1989),
- that reviewers narrow the scope of their reviews down to an explication and evaluation of one or two subjectively chosen design features of a dictionary, choice of lemmas being a favourite one (cf. Landau 1989: 305-306),
- that reviewers do not adequately motivate the choice of specific design features for evaluation, often focusing on what lexicographers themselves would perceive as trivial with regard to the processes of designing a dictionary or the design of the dictionaries themselves (Osselton 1989: 225),
- that reviewers often do not explicate their evaluation criteria, or indicate what their (lexicographical) status/validity are,[†] and
- that if reviewers make evaluative pronouncements on the design features of dictionaries, the criteria are often only implied and it is left to the reader to deduce them — a strategy whereby reviewers circumvent their task of providing a well-defined and well-motivated set of objective criteria for the evaluation of a dictionary (cf., for example, Bergenholtz and Mogensen 1994, Chan and Taylor 2001, Chapman 1977, Osselton 1989, Ripfel 1989, Rossenbeck 1994, and Steiner 1994).

Of course there are exceptions, but the overwhelming verdict therefore seems to be that reviews often lack objectivity, validity and reliability, and thus are in principle unfit to perform two of their major goals/functions: (1) to assist readers in their decision-making in acquiring the best dictionaries for their usage needs by presenting them with a well-founded analysis of the positive and negative qualities of a dictionary/dictionaries under review, and (2) to assist lexicographers in optimizing the functionality of their dictionaries.

Improving the quality of the praxis of dictionary criticism is a task that has been left to the scientific study of dictionary criticism/evaluation (*Wörterbuchkritik*) — one of the major components of a theory of lexicography. Although it is by no means clear precisely what the goals, methods and theories of this component of dictionary research should be (cf., however, Wiegand 1994), research in the field suggests that the scientific study of dictionary criticism should encompass the following:

- (1) to improve the quality of the evaluation of dictionaries and the advice reviewers have to give their target readers
 - (a) by providing clear and lexicographically relevant definitions of concepts such as 'evaluation/criticism' and 'positive/negative value' (cf. Ripfel 1989, and Wiegand 1994) and by specifying how to operationalize them,
 - (b) by providing reviewers with sets of explicitly formulated, generally acceptable, comprehensive, systematic and operationalizable criteria for the evaluation of dictionaries of all types,
 - (c) by providing generally acceptable methods to conduct such reviews (cf. Chapman 1977, Hartmann 2001, and Jackson 1996 for a discussion of approaches or methods),
 - (d) by providing the evaluation criteria in formats which are usable/functional for different methods for the review/evaluation of dictionaries of all types, and
 - (e) by providing a theory of the review as genre/document type (and specifying the competencies required from reviewers, goals, target readers and the design guidelines (content, structure, style, presentation, etc.) for the various document types in which reviewers do/could report on their evaluation of dictionaries (cf. Bergenholtz and Mogensen 1994, Chan and Taylor 2001, Jackson 1996, Ripfel 1989, and Rossenbeck 1994), and
- (2) to support lexicographers in the design of dictionaries of all types to optimally achieve the information and communication needs of their intended users in different contexts of use.

In this article, the focus falls on goal (1)(b), viz. that the scientific study of dictionary criticism has to provide reviewers with sets of explicitly formulated, valid, generally acceptable, comprehensive, systematic and operationalizable

criteria for the evaluation of dictionaries of all types. This is a goal which has neither been clearly stipulated nor achieved within the field of the scientific study of dictionary criticism. Working bottom-up from corpora of reviews, the main findings of such studies often only reflect the critique mentioned above, viz. that reviews focus on the design features of dictionaries and seldom evaluate them in terms of clearly specified criteria. As a result, such studies provide lists for the evaluation of dictionaries but these in fact consist for the larger part mostly of the design features of dictionaries which receive attention in reviews — the criteria in terms of which they have to be evaluated often receiving much less attention. Evaluation criteria are often only implicitly formulated, the key concepts are not distinctly defined, and it is not clear what status or validity the criteria have. It is seldom explicitly indicated how they are to be operationalized, and for what kind of reviewing method(s) they are to be used (cf., for example, Chapman 1977, Chan and Loong 1999, Chan and Taylor 2001, Jackson 1996, 2001, Jehle 1990, Kister 1992, Nakamoto 1994, Osselton 1989, Steiner 1984, 1994, and Wiegand 1994, 1998, 2002). Furthermore, with a few exceptions (cf., for example, Ripfel 1989), little effort has been made to explicate, analyse and evaluate the sets of evaluation criteria which are used in the reviews. Little thought has gone into such issues as to what the distinguishing features of evaluation criteria are, how best they can be systematized, in what formats they could best be presented for various kinds of reviewing activities, and how and according to which parameters these evaluation criteria are themselves to be evaluated.

Evaluation criteria for dictionaries are, however, not only presented in dictionary reviews — they also feature prominently and can be harvested from numerous other sources such as lexicographically relevant articles, manuals, handbooks, reports, etc. Although there are exceptions, much the same critique noted above applies to the way in which evaluation criteria are treated in these studies. Although some dictionary types, their design features and evaluation criteria have received more attention in the literature than others, one could hypothesize that for most dictionary types there still does not exist a set of explicitly formulated, valid, generally acceptable, comprehensive, systematic and operationalizable evaluation criteria.

1.1 Goals

The main goal of this article is to report on some groundwork towards the development of a lexicographically motivated framework for the description (analysis and classification) and evaluation of dictionary evaluation criteria. Working bottom-up from lexicographically relevant literature on dictionary evaluation criteria, a number of parameters for the description and evaluation of evaluation criteria will be proposed. Given the lack of research on this topic in the field of lexicography, however, literature on the criteria for the evaluation of documents, specifically Renkema's (1996) 3Cs model, and the frame-

work developed by De Jong and Van der Geest (2000) for the description and evaluation of website design heuristics were also consulted. Relying on usability research in the field of web design for this purpose is not arbitrary. Both the functional approach in the design and evaluation of dictionaries and the usability approach in the design and evaluation of websites (and documents in general) focus in their evaluation on the degree to which the design of the product (a dictionary, a website, a printed document) supports the effectiveness, efficiency and satisfaction with which different kinds of users can perform certain functions in various contexts of use.

In this article, owing to limitations of space, the development of the analytical framework for evaluation criteria is restricted to criteria for printed dictionaries. Although most of them also extend to the evaluation of electronic dictionaries, a number of additional criteria come into play in the case of electronic dictionaries, given the specifics of the medium, an increase in design options, and related usability issues (cf., for example, De Schryver 2003 for an analysis of the design features of electronic dictionaries).

2. A framework for the description and evaluation of evaluation criteria

Because of the lack of research on the parameters for a framework for the description and evaluation of dictionary evaluation criteria, a bottom-up approach is used in this section by incorporating the criteria used in Ripfel (1989) — the first of the more comprehensive studies of dictionary criticism (cf., however, also Jehle 1990, and Nakamoto 1994) — with a strongly modified version of the parameters for the analysis of website heuristics proposed by De Jong and Van der Geest (2000).

Ripfel (1989) describes and classifies dictionary evaluation criteria according to the following parameters:

- (1) the source of the evaluation criterion (journalistic reviews, expert reviews),
- (2) the dictionary/dictionary type to which the evaluation criterion pertains,
- (3) the specific design feature(s) of the dictionary type which the criterion addresses, and
- (4) the evaluation (positive vs. negative) attached to the criterion.

De Jong and Van der Geest (2000) cluster their features for the analysis of heuristics under four major headings, adjusted here to apply to dictionary evaluation criteria:

- (1) the information covered by the evaluation criteria,
- (2) the presentation format of the evaluation criteria,
- (3) the validity of the evaluation criteria, and
- (4) the application of the evaluation criteria.

<p>1. Information covered by the evaluation criteria</p> <p>(a) Specificity General Genre-specific Feature-specific</p> <p>(b) Exhaustiveness Exhaustive — Arbitrary</p> <p>2. Presentation format of the evaluation criteria</p> <p>Instructions Questions Requirements Explanations</p> <p>3. Validity of the evaluation criteria</p> <p>(a) Foundations Standards Theory User research Practitioners</p> <p>(b) Novelty value High — Low</p> <p>4. Application of the evaluation criteria</p> <p>(a) Level of expertise required Mechanistic — Expert</p> <p>(b) Outcome possibilities Open — Closed</p> <p>(c) Operationalization Specified — Unspecified</p>

Table 1: Parameters for the analysis and evaluation of dictionary evaluation criteria derived from De Jong and Van der Geest (2000)

By way of introduction, it seems necessary though to distinguish between statements which merely focus content-wise on the design features of dictionaries and those which are evaluation criteria, i.e. which also indicate what value should be attached to a design feature/design features and why. This is illustrated by means of Jackson's (1996) distinction between internal and external evaluation criteria.

2.1 Descriptive versus evaluative statements

Jackson (1996: 5-6) divides his proposed evaluation criteria for monolingual dictionaries into internal and external evaluation criteria:

Internal criteria derive from what a dictionary says about itself, or what the editors claim for the dictionary. External criteria derive from metalexigraphy, taking into account the linguistic requirements for a lexical description, as well as (*sic*) considerations of dictionary design and production.

As an example of internal criteria, Jackson (1996: 5) gives the claim about their defining strategy made by the compilers of the *New Oxford Dictionary of English* (NODE) in their introduction:

Each word has at least one core meaning, to which a number of subsenses may be attached ... Core meanings represent the typical, central uses of the word in question in modern standard English ... The core meaning is the one that represents the most literal sense the word has in ordinary modern usage ...

Jackson (1996: 5) adds to this quotation that 'these are testable statements', and, indeed, one could check a number of lexical entries and determine if their meanings are defined according to this defining strategy. However, in the quotation above, no evaluative statements about the efficacy of the defining strategy are given — the compilers merely describe how the meaning of words are explained and no claims of an evaluative kind are made, for example, that the defining strategy is better with regard to other defining strategies (or those of other dictionaries) in achieving some functional goal, such as full comprehension of the meaning of a lexical item. Clearly, statements about the design features of dictionaries should be distinguished from evaluative claims about the (in)efficacy of these features with regard to some goal or norm of efficacy (e.g. maximal comprehensibility).

Furthermore, if dictionaries deviate from their defining strategy in certain entries (should specific lexical items require it), it would hardly be considered a lexicographical sin; but in as much as it is an oversight, and occurs often in a dictionary, it would rather entail an infringement of the criterion of consistency (in applying design guidelines). What is at stake here then is an infringement of a high-level evaluation criterion (consistency) which holds for all design guidelines of a dictionary and which has its own motivation (cf. the discussion below).

A common trend in the analyses of dictionary reviews (cf., for example, Chan and Taylor 2001) and in the presentations of sets of evaluation criteria is, however, that the evaluation norms/dimensions (e.g. ease of comprehension of definitions in learner's dictionaries) that motivate a design feature (e.g. the use of a controlled/limited vocabulary in definitions) are simply omitted or only implied. Consequently, the mere absence or presence of descriptive features is then evaluated in themselves as either positive or negative without reference to the relevant motivating evaluation norm.

It should also be obvious from the example given above, that internal and external criteria are inadequately differentiated, and can in fact overlap: internal criteria can also be derived from metalexigraphy, or for that matter from linguistic requirements, or considerations of the design and production of a

dictionary (thus making them less useful parameters in terms of which to analyse evaluation criteria). The defining strategy outlined for NODE is actually followed by other dictionaries as well and not in itself seen as a negative practice.

Reviewers also commonly take the pronouncements made by compilers in the front matter of their dictionaries or those on their covers as a starting point for a review and then clearly link their analysis and evaluation to what Jackson would consider as 'external' criteria. A good example is Feinauer's (2007) review of the *Woordeboek van die Afrikaanse Taal Deel XII*.

2.2.1 Information covered by the evaluation criteria

2.2.1.1 Source of the evaluation criterion

Ripfel (1989) distinguishes between evaluation criteria used in journalistic reviews and those used in expert reviews. The genre of review types is expanded to five subtypes in Bergenholtz and Mogensen (1994) to include the following: extensive review for lexicographers, shorter review for lexicographers, extensive review for users (and lexicographers), shorter review for users (and lexicographers) and the book announcement. Given their differences in target readers, goals and length, one could expect different and more or less aspects of the design of a dictionary to be treated in more or less depth, and perhaps also differences in how explicitly reviewers will outline and motivate their evaluation criteria (cf. Bergenholtz and Mogensen 1994: 10, and Ripfel 1989: 31). However, it is not obvious whether different kinds of evaluation criteria are used in the different types of reviews or whether some of these are restricted to a certain type of review, making the source of evaluation criteria a less useful choice for differentiating between them.

Secondly, given that the same evaluation criteria can be used in any of these subtypes of reviews, it does not seem to be possible to differentiate between these evaluation criteria in terms of the subgenre of reviews in which they are used. However, one could assume that the author(s) of a review and/or the publication in which a review appears, and thus also the set of evaluation criteria used in the review, may affect their status. This issue needs more empirical research and until then the genre subtype or source does not seem to be a useful parameter for the description and evaluation of evaluation criteria other than being important for documentation/referencing and control purposes.

2.2.1.2 Scope of the evaluation criterion

Dictionary evaluation criteria differ in their scope. Parameter (2) (dictionary type) and parameter (3) (design feature) in Ripfel's set of parameters, and parameter (1) of De Jong and Van der Geest (2000) (information covered by the evaluation criteria) refer to the scope of an evaluation criterion. Some evalua-

tion criteria may hold for all dictionary types, others may be dictionary-specific; some evaluation criteria may hold for all the design features of dictionaries, others may only focus on one such design feature. A few examples are given in the following paragraphs.

2.2.1.3 Dictionary type

In his analysis of a corpus of bilingual dictionary reviews, Tomaszczyk (1988) indicates that only three (equivalents, directionality, reversibility) of the nine design features evaluated in his reviews (equivalents, directionality, reversibility, alphabetization, retrievability, redundancy, coverage, currency, reliability) pertain to the bilingual dictionary itself. Tomaszczyk (1988: 289) summarizes reviewers' evaluation criteria with regard to equivalents as follows:

The equivalents should be of an insertable kind, i.e. capable of being used in actual texts, and preferably monolexemic ... The equivalents proposed should be carefully selected closest possible ones rather than cross-linguistic (near) synonyms ... Definitions are allowed only with 'equivalentless' lexis ... Even these (i.e. definitions — PS) should be formulated in such a way (i.e. abbreviated) as to be substitutable ... In view of the fact that one-to-one lexical correspondences across languages are rare, use should be made of meaning (sense) discriminations. In a bidirectional dictionary, ... these should be given in the source language and for every discrete meaning in the target language ...

Obviously, it would be a task for research on dictionary typologies and dictionary structures — two major components of the theory of lexicography — to distinguish between the design features which are unique to a certain dictionary type and dictionary features which hold for all dictionaries, and a task for research on criticism to indicate what criteria are needed to evaluate all these features.

2.2.1.4 General evaluation criteria

Some evaluation criteria are applicable to dictionaries of all types. Good candidates for general or high-level criteria are Renkema's 3Cs model in which he proposes three general criteria, viz. correspondence, consistency and correctness. These evaluation criteria cut across all the other design guidelines and evaluation criteria and may have slightly different interpretations as they apply to each of these guidelines and criteria.

The one criterion most often reverted to in dictionary criticism is the criterion of consistency in the application of design guidelines. Dictionaries often deviate from their design guidelines, a common one being the lexical items they select from the general vocabulary for treatment in the dictionary. For example, De Schryver (2005) indicates at length how many dictionaries transgress this criterion by inconsistencies in the selection or omission of lexical items or in the over- or under-treatment of certain lemmas in dictionary arti-

cles. Other reviewers focus on such aspects as deviations in the ordering of lexical items in the wordlist of a dictionary and their microstructural treatment, or deviations in the defining strategy.

Tomaszczyk (1986: 293) links consistency to the *reliability* of the dictionary (which is increased if lexicographers adopt a set of clearly defined principles and adhere to them consistently). However, consistent application of the design guidelines of a dictionary ensures that users, once they get acquainted with the design of the dictionary, will be able to know exactly what information they can expect to be present in the dictionary and where to find it. In general, though, consistency in the application of the design guidelines leads to dictionaries having certain features, viz. 'predictability, calculability, analyzability and controllability' (Gouws and Prinsloo 2005: 9).

Correctness simply entails that whatever information the lexicographer provides to the user must be correct (although one might have to add the rider: 'in as far as the lexicographer could determine'). In contrast to correctness, correspondence may at first sight seem to be a vaguer criterion. However, Rundell (1999) addresses both these criteria in his deliberation on the improvement in the quality of the information provided in monolingual learner's dictionaries and the improvement in the way this information is presented to meet the needs and expectations of the potential users of a dictionary. More specifically he states that the correctness of a dictionary is optimized in as much as 'the description of a language that a dictionary provides, corresponds more closely to "the truth", that is, to reliable empirical evidence regarding the ways in which the language is used'. Obviously, the relevant empirical evidence referred to here is what one deduces from an acceptable corpus of the use of the language the dictionary purports to describe.

Correctness can of course extend beyond the grammatical information provided in a dictionary as it is a criterion that could be applied to all this information. Closer adherence to the correctness criterion is seen as positive; providing incorrect information is evaluated negatively. Rundell (1999: 88) does, however, relativise mere correctness (for example on the basis of a corpus analysis) against 'lexicographical relevance', i.e. the cluster of factors which enables a lexicographer to distinguish between information that is simply 'true' and information that is relevant to a specific dictionary (and type of user). For instance, *utter* can be used intransitively (as attested by examples in a corpus), but this is not relevant in a learner's dictionary.

Rundell (1999: 83) gives the criterion of correspondence a functional interpretation when he links it to the degree in which the presentation of the information provided in a dictionary corresponds 'more closely to what we know about the reference needs and reference skills of the target audience'. Correspondence is therefore the kind of high-level criterion tested in reviews in which reviewers try and determine how the design features of a dictionary do or do not support the information needs of different kinds of dictionary users with different linguistic competencies and reference skills in different contexts of language use (cf., for example, Bogaards 1996).

2.2.1.5 Feature-specific evaluation criteria

With regard to the third parameter, Ripfel (1989) structures her evaluation criteria according to an extensive set of 15 thematic categories which cover a variety of topics, including various aspects of the design features of the front and back matter and the central list of the reviewed dictionaries (cf. Ripfel 1989: 93-97 for a discussion). This practice links with a strategy used by some researchers (cf., for example, Bergenholtz and Mogensen 1994, Chan and Taylor 2001, Chan and Loong 1999, Jackson 1996, 2001, and Tomaszczyk 1986) in which the various evaluation criteria (focusing on only one design feature each) are organized according to the frame structure of dictionaries (or parts thereof): the front and back matter of a dictionary and their integration with the central wordlist, and various aspects of the design of the central wordlist, its macrostructure (choice of lemmas, lemmatization and outer and inner access structures), its microstructure (data categories and their structure), and the mediostructure (cf. Gouws and Prinsloo 2005 for a recent overview).

A good example is Jackson's (1996: 7-11) proposed criteria for the evaluation of monolingual dictionaries as presented in Table 2. His explication of the relevant criteria is also an example of the variety of ways in which evaluation criteria are presented for prospective reviewers and of how vague they in fact may be.

Content category	Evaluation criteria
Range of vocabulary	'A reviewer would need to determine whether, for its size and scope, the dictionary had adequate coverage of up-to-date, technical, international and, if appropriate, regional lexis.'
Word formation	'The judgment to be made is whether the account of word formation enables a user to ascertain the formal (morphological) relations between words.'
Homographs	No evaluation criterion. (Cf., however: 'The criteria for determining what is a headword have important consequences for lexical description as well as for accessibility.')
Sense division	'The issue here relates both to the adequacy of the lexical description and to how straightforward it is for the user to find the desired sense.'
Defining	'Not only does a reviewer need to assess the adequacy of the definitions, but also whether they are stylistically appropriate for the intended user.'
Beyond definition (lexical relations, collocations, connotations, etc.)	No evaluation criterion.
Pronunciation	'There are two issues here: the transcription system, ... and the accent to be represented.'

Grammar	'Learners' dictionaries have aimed for full coverage, which raises the question about how grammatical information is represented for effective access. A reviewer needs to evaluate how much information about the grammatical operation of words is necessary for a dictionary to fulfil its recording function, as against the need not to provide too much unnecessary information for the intended users.'
Usage	'The extent to which dictionaries are consistent in using their range of usage labels and how they apply them are matters for the critic to evaluate.'
Examples	'A number of questions need to be asked by the critic, relating to: the extent of the use of examples, what role they are seen to play in exemplification, where the examples come from (corpus or invented), and how consistently the dictionary's policy on examples is implemented.'
Etymology	'It is a matter of critical evaluation whether the information in the dictionary under review is appropriate to its size, purpose and intended users.'
Special features, e.g. synonym essays, boxed comments, usage notes, etc.	'The question is whether they add to the lexical description and the coverage and usefulness of the dictionary.'

Table 2: Criteria for the evaluation of monolingual dictionaries proposed by Jackson (1996)

It would be the task of research on dictionary typologies, dictionary structures and dictionary use to define a frame structure for each dictionary type. Much of this has already been done, for example, for the monolingual explanatory dictionary. Besides outlining relevant design guidelines and evaluation criteria for the front and back matter and the macrostructure of this dictionary type, much research focuses on the design of the microstructure. A number of design guidelines exist with regard to (1) what information categories should be included in the treatment of lemmas in the articles of general monolingual dictionaries (e.g. orthography, morphology, word class, meaning, idioms, etc.), (2) how this information should be encoded, and (3) how this data should be structured in the article as a whole or within the sections of the article dealing with specific information categories. (See in this regard the 'classical' data matrix provided in Hausmann and Wiegand (1989) and the linguistically motivated one given in Hudson (1988: 311-312).) Design guidelines for a variety of dictionary types and criteria to evaluate them are discussed extensively in Hausmann et al. (1989: 409-1056).

Dictionary-specific frame structures can be used as basis for the systematization of evaluation criteria for each dictionary type. These frame structures can also be used to identify for which aspects of the design of a dictionary type no relevant or inadequate evaluation criteria have been proposed and to meas-

ure the comprehensiveness and systematicity of the proposed evaluation criteria. (For the use of dictionary functions as a basis for organizing the design features of dictionaries and their associated evaluation criteria, see the discussion below.)

2.2.1.6 Exhaustiveness

The parameter of comprehensiveness ties in with De Jong and Van der Geest's (2000) parameter of exhaustiveness, which refers to how well an evaluation criterion or a set of evaluation criteria cover a design feature/features of a dictionary. If one compares, for example, Jackson's (1996) two criteria for dictionary definitions above (adequacy, appropriate style) with McMillan's (1949) (i.e. dictionary definitions should be complete, clear, accurate, consistent, independent, objective, and neutral), it is obvious that the former are not as exhaustive as the latter (cf. Gouws and Prinsloo 2005: 147-148 for an explanation of each of these criteria).

On the other hand, McMillan's (1949) framework for the evaluation of college dictionaries is more exhaustive, but obviously does not cover every aspect of the design of such dictionaries as can be verified against the frame structures proposed for monolingual explanatory dictionaries (cf. references above). McMillan's criteria relate to the following aspects of the design of such dictionaries (cf. Mdee 2004: 370-371):

- (1) The quantity of the information the dictionary offers. This includes the number of entries, meanings, new words (when compared to competitors), synonyms, pronunciation, etymology and the use of subject and usage labels.
- (2) The quality of the information presented in the dictionary. The aspects to examine here are:
 - (a) the accuracy of the information,
 - (b) the completeness of the information,
 - (c) the clearness of the information,
 - (d) the simplicity in presenting the information, i.e. the information should be rendered in such a way that it can easily be deciphered, and
 - (e) the modernity of the dictionary, i.e. the number of current words in the language which has been included in the dictionary.
- (3) The effectiveness of the presentation of the information. This includes the order of arranging entries, the placement of etymology, the ordering of senses, and the presentation of pronunciation. Also important is the typography.

As a dictionary type, learner's dictionaries have been extensively reviewed with regard to most of their relevant design features, such as the lists provided by Bogaards (1996) and Ilson (1999). Dretzke (1997) has compiled a list of thirteen evaluation criteria, mostly derived from design guidelines for the content categories, to aid students and teachers to assess the value of a learner's dic-

tionary. A much more exhaustive list, however, is provided by Chan and Loong (1999) — a fact that underlines the relativity of a concept such as exhaustiveness when it is measured in terms of the sets of evaluation criteria that have been developed.

2.2.1.7 Value

Ripfel (1989) lists both the positive and negative evaluation criteria for each aspect of the design of the reviewed dictionaries (cf. Ripfel 1989: 137 ff.). For example, one of the positive criteria with regard to the quality of the meaning explanations of a dictionary is also the provision of encyclopedic information ('die Bedeutungserläuterungen geben auch Auskunft über die Sache selbst, der das Wort gilt'); a negative criterion/tendency is the omission of some meaning distinctions (cf. Ripfel 1989: 153).

Both these evaluation criteria are, however, problematic in as much as it is not made clear why they are considered to be positive or negative. For example, the omission of meanings has a negative effect on users finding the semantic information they need. The relevant functional evaluation criterion in this case is that for reception purposes, users must be able to easily find the meaning of the word they may look up (cf. Bogaards 1996 and the discussion below). Providing as many meanings as possible for each lemma furthers this goal; providing less, hampers it.

The status of the positive evaluative statement provided as example is also uncertain. It is based on a distinction between encyclopedic meaning and linguistic meaning, which is difficult to uphold. However, given that such a distinction could be made, it is not clearly stated why it would necessarily be a positive feature if a meaning explanation in a dictionary includes encyclopedic information. Unless this motivation is spelt out, it is rather problematic to assess the validity and usefulness of such a criterion.

Whether it is in fact necessary or useful to make a distinction between positive and negative evaluative statements or evaluation criteria is not clear. Any evaluation criterion can implicitly be seen as being positive (in as much as it is motivated; cf. the discussion below) and any deviation from it as negative to a certain degree. However, as will be discussed below, further empirical research is necessary on the various forms in which evaluation criteria are or could be formulated before this distinction is simply abandoned.

2.2.2 The presentation format of dictionary evaluation criteria

As should be evident from the examples provided so far, evaluation criteria may be formulated in a variety of ways, only some of which are indicated in De Jong and Van der Geest (2000) (cf. Table 1 above): instructions, questions, requirements or explanations. Evaluation criteria are often only implicitly stated and first have to be made explicit in any one of the various formats above.

Some evaluation criteria may be formulated in any one of these formats and a specific format may be necessitated by the function for which the criteria are to be used. For example, McMillan's (1949) criteria for definitions may be formulated as an instruction/requirement ('Definitions must be/have to be complete, clear, accurate, consistent, independent, objective, and neutral.') or as a question ('Are the definitions complete, clear, accurate, consistent, independent, objective, and neutral?').

In this regard, De Jong and Van der Geest (2000) differentiate between process-orientated heuristics and product-orientated heuristics. In lexicographical terms, such a distinction would hold when evaluation criteria are used as design guidelines in the process of compiling a dictionary, typically formulated as instructions in a plan of the dictionary (cf. Gouws and Prinsloo 2005: 9-19), or as criteria for evaluating the dictionary as a finished product, typically formulated as a set of questions in a checklist. A good example is the checklist provided by Chan and Loong (1999) for learner's dictionaries. Little research has been forthcoming, however, on what formats would best suit what kind of function for which evaluation criteria used.

2.2.3 The validity of evaluation criteria

Dictionary reviewers have been criticized for the fact that they evaluate dictionaries in terms of subjective considerations and seldom motivate their choice of evaluation criteria. When evaluating a dictionary, it is therefore important to know how valid the criteria are on which such an evaluation is based. Ripfel (1989) also touches on the importance of distinguishing between valid evaluative statements and those merely representing the personal tastes of a reviewer. As will be indicated below, it is necessary to distinguish between the motivation/validation for a design feature, the criterion that targets it and the value (positive or negative, be it on a scale) that one should associate with such a criterion.

In the rest of this section, the motivation/validation for dictionary evaluation criteria are discussed in more detail.

2.2.3.1 Dictionary research

Given that dictionary criticism as a field of dictionary research is dependent on the other components of lexicographical research, the design features of dictionaries and their associated evaluation criteria may find their validity in or be motivated by any of the other areas of dictionary research. These areas include

- research on the history of dictionaries (which could explain/motivate why dictionaries have certain features in response to their changing social-cultural contexts),
- research on dictionary typologies and the distinctive features of dictionary types,

- research on dictionary use and users such as the linguistic competencies of target users, their reference skills, and their needs in different contexts of use (cf. the functional validity of evaluation criteria discussed below), and
- research on dictionary structure.

As has already been indicated in the discussion above, each component of the theory of lexicography feeds into the description, evaluation and development of dictionary evaluation criteria in various ways. However, with regard to the validity of evaluation criteria, the functional motivation of the design of dictionaries is currently strongly in focus.

2.2.3.1.1 Functional motivation

In recent years, the functional approach has dominated in the field of dictionary design and evaluation (cf. Bergenholtz and Tarp 2003, Nesi 1999, Swanepoel 2001, Tarp 2007, and Wiegand 1998). According to this approach, each element of the design of dictionaries should be determined/motivated by (1) the linguistic competencies and reference skills of the target users of these dictionaries, and (2) the functions dictionaries have to fulfil in various contexts of use (cf. cognitive and communicative functions discussed in Bergenholtz and Tarp 2003, and Tarp 2007), within the confines of their typological features, and subject to the real world constraints under which they are produced.

In as much as the design features of a dictionary successfully supports the target users and their linguistic needs in such use contexts, they are evaluated positively; in as much as they do not, they are evaluated negatively. Degrees of functionality/usability on the positive–negative scale is thus defined in terms of the performance levels achieved by users for certain tasks in which a dictionary with specific design features is used. As such, it accords with the user approach to the evaluation of a dictionary.

This approach differs from that of the expert checklist in which the functionality/usability of a dictionary is defined in terms of the absence (negatively evaluated) or presence (positively evaluated) of certain design features. As Dillon (2001) points out, however, such an approach assumes that functionality can be an inherent feature of a dictionary if it has certain design features. However, for any combination of target users, specific tasks and usage contexts, there could always be certain aspects of the design of a dictionary which might not function optimally. Furthermore, as will be indicated below in a discussion of Bogaards's (1996) evaluation of a set of learner's dictionaries, experts often have to add that they cannot make a final evaluation of the functionality of a design feature of a dictionary before its success (or not) has been empirically tested with users.

A good example of the way dictionaries can be evaluated by determining how their design features do/do not to a degree support their target users with

their reference skills and specific linguistic needs in certain contexts of use is Bogaards's (1996) review of four learner's dictionaries (*Collins COBUILD English Dictionary*, *Longman Dictionary of Contemporary English*, *Oxford Advanced Learner's Dictionary of Current English*, and *Cambridge International Dictionary of English*). Bogaards (1996) spells out the reference skills and linguistic competencies one can assume the target users possess and the design features needed to optimize the usability of the dictionary for them in specific contexts of L2 acquisition. For example, for receptive purposes — on which the focus is here — finding the meanings of problematic words from texts in the dictionary (the findability problem) and understanding the explanations provided of their meaning (the comprehension problem) are critical. Both of these require that dictionaries should have a number of design features to optimize the findability of the relevant meaning(s) and the comprehensibility of the explanations provided (functional motivation). Furthermore, Bogaards explicates what data categories for lexical items have to be provided in a dictionary for a user to acquire a lexical item fully for both receptive and productive uses of such items (linguistic motivation/lexical acquisition).

Bogaards (1996) discusses in more detail what design strategies the various dictionaries use to address the findability and the comprehensibility problem and, in most cases, to what extent they can or could do so successfully or not, or at least to what degree, relative to each other or pending empirical testing with users. The relevant reception design strategies/features, plus the relevant analytical questions Bogaards (1996: 280) provides are summarized in the following table (Bogaards 1996: 315):

Findability	<ol style="list-style-type: none"> 1. Number of meanings explained (How many words and expressions are entered?) <ul style="list-style-type: none"> – morphological tools (to decipher words not in the dictionary) 2. Accessibility of forms (Are all word forms easily accessible?) <ul style="list-style-type: none"> – meaning-related forms – irregular forms 3. Accessibility of multiword expressions (Where can expressions be found?) 4. Structure of entries (What is done to guide the user in longer entries?) <ul style="list-style-type: none"> – use of labels – guiding principles – general layout
Comprehensibility	<ol style="list-style-type: none"> 5. Definitions (How comprehensible are the definitions given?) <ul style="list-style-type: none"> – defining vocabulary – precision – defining style 6. Illustrations, etc. (What types of illustrations or other devices are used to make meanings clear?) <ul style="list-style-type: none"> – pictures

	<ul style="list-style-type: none"> - synonyms, etc. - notes, etc.
7. Examples (How comprehensible are the examples given?)	

Table 3: Reception design strategies/features and analytical questions used by dictionaries, according to Bogaards (1996)

As the relevant evaluation criteria are not spelt out, such a list, however, does not differ from those that merely indicate what aspects of the design of a dictionary should be described in its evaluation. One cannot simply assume, as is often done in usability studies, that the mere presence of these design features would ensure or optimize findability or comprehensibility.

Sometimes the relevant evaluation criteria are easy to formulate. For example, with regard to the ordering of multiword expressions, Bogaards (1996: 286) notes that users with different mother tongues have different search strategies (motivation), and seeing that learner's dictionaries are aimed at an international market (motivation), 'they should ideally mention all multiword expressions in the entries of all relevant content words'. Bogaards (1996) is, however, also well aware of the fact that it is no simple task to come up with explicit and clearly formulated evaluation criteria with regard to all the relevant design features which are supposed to address the findability and comprehensibility problem. For example, with regard to parameter 1 in Table 3, the argument (and derived evaluation criterion) is that if more words are selected for treatment in a dictionary and if more meanings of words are explained, the chances are greater that the user will find the relevant meaning of a word. Bogaards (1996: 282) phrases the criterion as follows: 'As far as *findability for receptive* purposes is concerned, it is clear that the more lexical units there are in a dictionary, the better are the chances that a learner will find what he needs.' More would therefore be evaluated as positive, fewer as less positive.

In this case, however, what constitutes more and fewer is not stipulated independently or in absolute terms — more and fewer are relativised to the number of lemmas treated in the other dictionaries of the same type which are subject to the review. As Bogaards (1996: 282) notes, the interpretation of whether more and fewer should simply be seen in positive/negative terms may not be the correct approach: 'It is difficult to say whether or to what extent the selection of lexical units by one dictionary is better adapted to the needs of the learners than the choice made by another one.' Application of this criterion is also no easy task as, given the differences in design of various dictionaries, it is rather difficult to create a uniformly defined measurement instrument to count treated lemmas.

The following can serve as a second example. If one assumes that target users know the alphabet, the findability of the word form would be optimized if all forms are treated as lemmas, each with its own entries, and a strict vertical alphabetical ordering of all the lemmas are followed. In as much as the macro-ordering design features of a dictionary thus match this strategy, the more

positive it should be evaluated. Bogaards (1996: 284) spells it out: 'The simplest way of finding a word is to look it up in the alphabetical list. Whenever a form is not in its exact alphabetical place, the learner will have to step back from his text and wonder where he might find what he is looking for.' Nesting and niching strategies, including all multiword expressions in an outer text, and complex systems of cross-referencing worsen the findability problem and should therefore be viewed less positively. They are, however, strategies which most printed dictionaries have to follow given space constraints, and thus are not so easy to simply evaluate negatively.

In most cases, an expert reviewer can only indicate which of these design strategies seem to be or could be effective in addressing these two problems, but independent empirical testing with the target groups is necessary to back up the claims reviewers might make about the efficacy of such design features. With regard to the various definition styles used, Bogaards (1996: 293) clearly states:

In spite of the criticisms formulated, it is difficult to say anything definitive about which type of definition is most profitable for L2 learners. Very little research has been done on this topic. Indeed, whatever research there has been done leads to the conclusion that the dictionary does not seem to contribute in a significant way to a better understanding of a written text.

2.2.3.2 Standards

Standards issued by some authoritative organization would represent, ideally, the generally agreed upon and compulsory criteria for the design and evaluation of dictionaries. One such standard is the revised ISO 1951, which specifies a number of general guidelines for the design of dictionaries and specifically for the content categories and their structuring (cf. Le Meur and Derouin 2005). Such design guidelines could, for example, be used to assess to what extent dictionaries of a specific type in fact include the specified content categories and adhere to other prescribed design guidelines. However, adhering to prescribed guidelines in standards does not *per se* guarantee the efficacy of a dictionary. For example, adhering to the guidelines for the structuring of the data may, but will not necessarily, help users actually find the data they require. Whether or not it facilitates the findability of the data will also depend on whether or not the guidelines themselves have been derived from user research on findability and the structuring of information in dictionaries.

One of the major goals of most standards is to foster uniformity and thus compatibility. However, improving the quality of the product — here, all types of dictionaries — may also be a major goal, but then one which could only be attained if the standards as design guidelines or as evaluation criteria are well motivated by research on product/dictionary use and product design/dictionary structure.

2.2.3.3 Theoretical motivation

Evaluation criteria may also be motivated by lexicographical or any other relevant theory, for example, learning and information-processing theory or theories of language production and comprehension (cf., for example, Van de Poel and Swanepoel 2003). Linguistic theories have by and large dominated as motivation for the content and structuring of the grammatical information in dictionaries. Structuralism has played a major role in this regard (cf., for example, Zgusta 1971), but numerous other theories have had an influence on both the selection and the presentation of grammatical information, such as theories of the mental lexicon, theories of semantic networks, frame semantics and cognitive semantics (cf. Geeraerts 2001 and the articles on 'Homonymy and Polysemy' in the thematic part of *Lexicographica, International Annual for Lexicography*, 17: 1-181, 2001).

One of the major influences to come to the fore with the development of corpora and their use in dictionary compilation is the theory of language learning based on the principle of idiomaticity/phraseology/multiword expressions. This has led to a positive increase in the presentation of usage information for multilexical units in monolingual learner's dictionaries (cf. Béjoint 2000: 209-225; the articles on 'The Corpus Approach to Lexicography' in the thematic part of *Lexicographica, International Annual for Lexicography*, 20: 1-129, 2004 and the articles on 'Idioms in Mono- and Bilingual Lexicography' in the thematic part of *Lexicographica, International Annual for Lexicography*, 19: 1-139, 2003).

Certain design features may in fact have both a theoretical and functional explanation. Bogaards (1996), for example, specifies what (grammatical) information has to be included in the entries of learner's dictionaries for a target user to acquire enough lexical knowledge to use these lexical items in reception and production tasks.

2.2.3.4 Lexicographical practice as motivation

There may also be design features of dictionaries which are simply based on lexicographical practice (practice-based heuristics), many of which are conventional features of dictionaries for which no clear motivation may in fact exist.

2.2.3.5 Novelty as motivation

One of the major features of the so-called 'big five' learner's dictionaries is the innovative way in which they try to address a number of the problems that learners have in finding and comprehending the information they seek (cf. Bogaards 1996, Rundell 1999, and Swanepoel 2000). The novelty value of a design feature (as motivation for its use) may lie in the fact that it addresses a general user problem in a new way. However, novelty in design is not *per se* positive. As Swanepoel (2000, 2001) argues, very few of these design strategies have been tested empirically for their efficacy/functionality.

The foregoing discussion necessitates one to draw a distinction between the motivation for a design feature of a dictionary (e.g. a functional consideration relating to the target user or the context of use) and the evaluation of such a feature, i.e. the degree to which it successfully assists the user with a certain task in a specific content of use. For example, the use of a limited vocabulary in definitions in monolingual learner's dictionaries is motivated by the restricted linguistic competencies of its L2/L3 target users. Its functional goal (on which it has to be evaluated) is to enable better comprehension of the definitions provided. However, whether or not such a controlled vocabulary in fact leads to a clear comprehension of the meaning/sense which such a definition aims to impart is another question and subject to empirical testing with a sample of the target users of the dictionary. For example, such a vocabulary is sometimes too restrictive to explain subtle nuances of meaning of polysemous lexical items or meaning differences between related lexical items (synonyms). Although the question of the motivation for a criterion and its evaluative dimension may in fact overlap, it may be necessary to discern them in other cases.

2.2.4 Application of the evaluation criteria

Parameter 4 of De Jong and Van der Geest (2000) (see Table 1) focuses on the way in which criteria can be used for the evaluation of a dictionary. Subparameter 4(a) would refer to the kind of expertise required to determine whether or not a design feature of a dictionary adheres to the criteria. For example, it can be determined almost mechanistically whether information is provided on the pronunciation of a word, but it might be much more difficult or require greater expertise to determine whether or not the definitions in a dictionary adhere to McMillan's (1949) criteria (cf. the discussion above). Likewise, there may be only one or two possible ways in which information on pronunciation could be provided (a closed set), but numerous ways in which Bogaards's (1996) findability and comprehensibility problems may be addressed.

Subparameter 4(c) refers to whether or not the criterion specifies precisely how (or in what respects) it is to be applied or operationalized to come to an evaluation. Some evaluation criteria may require only some matching operation, others are more difficult to use. For example, how does one determine whether the information in a dictionary is easy to find? What would be expected though, is that all evaluative concepts in terms of which a criterion is formulated will be clearly defined and that it should be clearly indicated how to measure them. This is certainly no easy task. As far as could be ascertained, however, no research has been forthcoming on how reviewers go about evaluating dictionaries in practice.

3. Conclusion

From the discussion above, it seems necessary to distinguish between pure

descriptive statements about the design features of dictionaries and evaluation criteria clearly setting out the basis for the evaluation.

For the analysis of dictionary evaluation criteria a framework consisting of the following parameters is proposed here:

<p>1. Information covered by the evaluation criteria</p> <p>(a) Source</p> <p>(b) Scope of the evaluation criteria Dictionary type General evaluation criteria Feature-specific evaluation criteria</p> <p>(c) Exhaustiveness</p> <p>(d) Value</p> <p>2. Presentation format of the evaluation criteria</p> <p>(a) Formulation of items Instructions Questions Requirements Explanations</p> <p>(b) Process vs. product criteria</p> <p>3. Validity of the evaluation criteria</p> <p>(a) Foundations Dictionary research Functional motivation</p> <p>(b) Standards Theoretical motivation Lexicographical practice as motivation Novelty value as motivation</p> <p>4. Application of the evaluation criteria</p> <p>(a) Level of expertise required Mechanistic — Expert</p> <p>(b) Outcome possibilities Open — Closed</p> <p>(c) Operationalization Specified — Unspecified</p>

Table 4: Parameters for the analysis and evaluation of dictionary evaluation criteria

Although the framework presented in Table 4 is mainly meant for the analysis of dictionary evaluation criteria, one could propose, though, that to be usable, the evaluation criteria themselves will have to meet the following evaluation criteria: be explicitly formulated, valid/motivated, generally acceptable, and the evaluative concepts on which they are based will have to be clearly defined and operationalized — features discussed above.

The use of 'towards' in the title of this article is meant to capture the fact that the proposed framework for the description and evaluation of dictionary evaluation criteria is tentative. It is a first step towards the formulation of such a framework, which itself will no doubt change as the empirical base for this study is expanded and more and more studies on dictionary criticism are included.

Endnote

- † The term *evaluation criterion* refers in reviewing practice to any standard, norm, principle, rule of thumb, heuristic or procedure on the basis of which a design feature of a dictionary is evaluated. The term *design feature* itself refers to the 'building blocks' of dictionaries of various types, in particular the frame features of dictionaries, which includes their front and back matter/texts, the macro-, micro-, medio-, and access structures of the central wordlist, the interaction between these outer texts and the central wordlist, and the smaller elements of which these major structural components of printed and electronic dictionaries are constituted.

Bibliography

- Béjoint, H.** 2000. *Modern Lexicography. An Introduction*. Oxford: Oxford University Press.
- Bergenholtz, H. and J.E. Mogensen.** 1994. Wörterbuchkritik in Dänemark. *Lexicographica* 9: 8-35.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes* 31: 171-196.
- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*. Amsterdam: John Benjamins.
- Bogaards, P.** 1996. Dictionaries for Learners of English. *International Journal of Lexicography* 9(4): 277-320.
- Chan, A.C and Y. Loong.** 1999. Establishing Criteria for Evaluating a Learner's Dictionary. Berry, R., B. Asker, K. Hyland and M. Lam (Eds.). 1999. *Language Analysis, Description and Pedagogy*: 298-307. Hong Kong: The Hong Kong University of Science and Technology.
- Chan, A.Y.W. and A. Taylor.** 2001. Evaluating Learner Dictionaries: What the Reviews Say. *International Journal of Lexicography* 14(3): 163-180.
- Chapman, R.L.** 1977. Dictionary Reviews and Reviewing: 1900-1975. Raymond, J.C. and I.W. Russell (Eds.). 1977. *James B. McMillan: Essays in Linguistics by his Friends and Colleagues*: 143-161. Alabama: The University of Alabama Press.
- De Jong, M. and T. van der Geest.** 2000. Characterizing Web Heuristics. *Technical Communication* 47(3): 312-325.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- De Schryver, G.-M.** 2005. Concurrent Over- and Under-Treatment in Dictionaries — The *Woordeboek van die Afrikaanse Taal* as a Case in Point. *International Journal of Lexicography* 18(1): 47-75.

- Dillon, A.** 2001. Evaluation of Software Usability. Karwowski, W. (Ed.). 2001. *International Encyclopedia of Ergonomics and Human Factors*: 1110-1112. London/New York: Taylor and Francis.
- Dretzke, B.** 1997. Moderne englisch-englische Lernerwörterbücher: Beurteilungskriterien. *Fremdsprachenunterricht* 41(3): 218-223.
- Feinauer, I.** 2007. Doen die *Woordeboek van die Afrikaanse Taal Deel XII* dit vir die Afrikaanse taal en die Suid-Afrikaanse leksikografie? *Lexikos* 17: 258-277.
- Geeraerts, D.** 2001. The Definitional Practice of Dictionaries and the Cognitive Semantic Conception of Polysemy. *Lexicographica* 17: 6-21.
- Gouws, H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Hartmann, R.R.K.** 1996. Lexicography as an Applied Linguistic Discipline. Hartmann, R.R.K. (Ed.). 1996. *Solving Language Problems*: 230-244. Exeter: University of Exeter Press.
- Hartmann, R.R.K.** 2001. Methods in Dictionary Research. Hartmann, R.R.K. (Ed.). 2003. *Lexicography. Critical Concepts. Volume 3*: 429-444. London/New York: Routledge.
- Hartmann, R.R.K. (Ed.).** 1999. *Dictionaries in Language Learning. Recommendations, National Reports and Thematic Reports from the TNP Sub-project 9: Dictionaries*. Berlin Free University. <http://web.fu-berlin.de/elc/tnp1/SP9dossier.doc> [3 May 2008].
- Hausmann, F.J. and H.E. Wiegand.** 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J. et al. (Eds.). 1989-1991: 328-359.
- Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*. Berlin: Walter de Gruyter.
- Hudson, R.** 1988. The Linguistic Foundations for Lexical Research and Dictionary Design. *International Journal of Lexicography* 1(4): 287-312.
- Ilsou, R.** 1999. Nine Learners' Dictionaries. *International Journal of Lexicography* 12(3): 223-237.
- Jackson, H.** 1996. *Dictionary Criticism*. Unpublished manuscript. Birmingham City University, Faculty of Computing and Information Studies, Research Papers 2-8.
- Jackson, H.** 2002. *Lexicography: An Introduction*. London/New York: Routledge.
- Jehle, G.** 1990. *Das englische und französische Lernerwörterbücher in der Rezension. Theorie und Praxis der Wörterbuchkritik*. *Lexicographica. Series Maior* 30. Tübingen: Max Niemeyer.
- Kister, K.** 1992. *Kister's Best Dictionaries for Adults and Young People: A Comparative Guide*. Phoenix: Oryx.
- Landau, S.I.** 1989. *Dictionaries. The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Le Meur, A. and M.-J. Derouin.** 2005. ISO 195: A Revised Standard for Lexicography. *Kernerman Dictionary News* 14. <http://kdictionary.com/kdn/kdn14/kdn1401-iso1951.html> [30 April 2008].
- McMillan, J.B.** 1949. Five College Dictionaries. *College English* 10(4): 214-221.
- Mdee, J.** 2004. *Kirkeby's English-Swahili Dictionary*. *Lexikos* 14: 369-395.
- Nakamoto, K.** 1994. *Establishing Criteria for Dictionary Criticism: A Checklist for Reviewers of Monolingual English Learner's Dictionaries*. Unpublished M.A. Thesis. Exeter: University of Exeter.
- Nesi, H.** 1999. The Specification of Dictionary Reference Skills in Higher Education. Hartmann, R.R.K. (Ed.). 1999: 53-67.

- Osselton, N.E.** 1989. The History of Academic Dictionary Criticism with Reference to Major Dictionaries. Hausmann, F.J. et al. (Eds.). 1989: 225-230.
- Renkema, J.** 1996. Over smaak valt goed te twisten. *Taalbeheersing* 18(4): 324-338.
- Ripfel, M.** 1989. *Wörterbuchkritik. Eine empirische Analyse von Wörterbuchrezensionen*. Lexicographica. Series Maior 29. Tübingen: Max Niemeyer.
- Rossenbeck, K.** 1994. Wörterbuchkritik in Schweden nach 1945: Die zweisprachigen Wörterbücher. *Lexicographica* 9: 88-107.
- Rundell, M.** 1999. Thematic Report 6. Recent Trends in Publishing Monolingual Learner's Dictionaries. Hartmann, R.R.K. (Ed.). 1999: 83-98.
- Steiner, R.** 1984. Guidelines for Reviewers of Bilingual Dictionaries. *Dictionaries* 6: 166-181.
- Steiner, R.** 1994. Reviews of Dictionaries in Learned Journals in the United States. *Lexicographica* 9: 159-173
- Swanepoel, P.** 2000. Providing Lexicographic Support for SL Vocabulary Acquisition: What Kind, under What Conditions, for Whom and Why? Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 403-418. Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.
- Swanepoel, P.** 2001. Dictionary Quality and Dictionary Design: A Methodology for Improving the Functional Quality of Dictionaries. *Lexikos* 11: 160-190.
- Tarp, S.** 2007. Lexicography in the Information Age. *Lexikos* 17: 170-179.
- Tomaszczyk, J.** 1988. The Bilingual Dictionary under Review. Snell-Hornby, M. (Ed.). 1988. *ZüriLEX '86 Proceedings. Papers Read at the EURALEX International Congress, University of Zürich, 9–14 September 1986*: 289-297. Tübingen: Francke Verlag.
- Van de Poel, K. and P. Swanepoel.** 2003. Theoretical and Methodological Pluralism in Designing Effective Lexical Support for CALL. *Computer Assisted Language Learning* 16(2–3): 174-212.
- Wiegand, H.E.** 1994. Zur Einführung. *Lexicographica* 9: 1-7.
- Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexicographie. 1. Teilband*. Berlin/New York: Walter de Gruyter
- Wiegand, H.E. (Ed.).** 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II*. Tübingen: Max Niemeyer.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague/Paris: Mouton.