The Effect of Lexicographical Information Costs on Dictionary Making and Use*

Sandro Nielsen, *Centre for Lexicography, Aarhus School of Business, University of Aarhus, Aarhus, Denmark (sn@asb.dk)*

Abstract: An important aspect for any lexicographer to consider is the ease with which users will be able to acquire the necessary information from the data presented in the dictionary, whether electronic or printed. Furthermore, this aspect has high priority with users and reviewers. A distinction is proposed between two general types of lexicographical information costs. Firstly, search-related costs are the effort required by the look-up activities users have to perform when consulting a dictionary to find access to the data they are searching for. It is argued that the access route, article structure, data distribution and cross-references may influence search-related information costs positively as well as negatively. Secondly, comprehension-related costs are the effort connected to the user's ability to understand and interpret the data presented in a dictionary. In other words: How easy or difficult is it for users to understand the data presented? Examples show how textual condensation, dictionary functions and use-situations may impact on the level of comprehension-related information costs. It is thus possible to establish a framework for evaluating lexicographical information costs for the purpose of dictionary making, use, analysis and research.

Keywords: LEXICOGRAPHY, DICTIONARY, INFORMATION COSTS, ACCESS ROUTES, DISTRIBUTION STRUCTURE, CROSS-REFERENCES, TEXTUAL CONDENSATION, DICTIONARY ARTICLES, SEARCH ZONES, DATA, INFORMATION, DICTIONARY FUNCTIONS, USE-SITUATIONS

Opsomming: Die uitwerking van leksikografiese inligtingskoste op die maak en gebruik van woordeboeke. 'n Belangrike aspek vir enige leksikograaf om in ag te neem is die gemak waarmee gebruikers in staat sal wees om die nodige inligting van die data in die woordeboek, hetsy elektronies of gedruk, te verkry. Verder geniet hierdie aspek hoë voorkeur by gebruikers en resensente. 'n Onderskeid word voorgestel tussen twee algemene soorte leksiko-grafiese inligtingskoste. Eerstens, soekverwante koste is die inspanning wat gepaardgaan met die opsoekaktiwiteite wat gebruikers moet verrig by die raadpleging van 'n woordeboek om toegang te verkry tot die data waarna hulle soek. Daar word geredeneer dat die toegangsroete, artikel-struktuur, dataverspreiding en kruisverwysings soekverwante inligtingskoste sowel positief as negatief kan beïnvloed. Tweedens, begripsverwante koste is die inspanning verbonde aan die gebruiker se vermoë om die gegewe data in 'n woordeboek te verstaan en te interpreteer. Met ander woorde: Hoe maklik of moeilik is dit vir gebruikers om die gegewe data te verstaan? Voor-

---

* This article is partially based on a presentation made at the Thirteenth International Symposium on Lexicography at the University of Copenhagen, Copenhagen, Denmark, 3–5 May 2007.

beelde toon hoe tekstuele verdigting, woordeboekfunksies en gebruiksituasies 'n uitwerking kan hê op die vlak van begripsverwante inligtingskoste. Dit is derhalwe moontlik om 'n raamwerk te skep vir die beoordeling van leksikografiese inligtingskoste vir die doel van die maak, gebruik, ontleding en ondersoek van woordeboeke.

Sleutelwoorde: LEKSIKOGRAFIE, WOORDEBOEK, INLIGTINGSKOSTE, TOEGANGS-ROETES, VERSPREIDINGSTRUKTUUR, KRUISVERWYSINGS, TEKSVERDIGTING, WOORDE-BOEKARTIKELS, SOEKSONES, DATA, INLIGTING, WOORDEBOEKFUNKSIES, GEBRUIK-SITUASIES

1.   Introduction

When we consult dictionaries we are in effect seeking knowledge, but acquiring knowledge involves costs, whether we are aware of it or not. The costs involved in searching for data and in reading texts will always be present but are often so insignificant that they go unnoticed. Authors may have adapted their writings to a particular audience and to specific types of reading situations. It is not uncommon to find children's editions of books originally written for adult readers, which have been adapted for children in a language they can understand, for instance the *Encyclopedia Britannica 2007 Children's Encyclopedia.* The publishers have realized that children up to a certain age will not be able to read and understand the adult edition of the encyclopedia properly; children do not consult the *Encyclopedia Britannica* for adults because it is too difficult to read and understand.

We have all been in a situation where, after having consulted a dictionary, we feel let down because the dictionary did not provide the expected help. One of the reasons for our unhappiness with the result of our consultation may be that we did not acquire the information we hoped we would gain by looking up a word in the dictionary. The lack of knowledge development is usually caused by a range of factors, and it is important to be aware of these when working with dictionaries. In this article, I will look at some of the factors that may frustrate the process of consulting dictionaries. This involves an examination of the communicative role of dictionaries and the laying of a theoretical foundation of the relation between their data and the information users may acquire. I will then address factors that may lead lexicographers to choose more carefully between several options when making relevant decisions for the planning and compilation of a dictionary.

2.   Dictionaries in the communication process

Dictionaries may be regarded as elements in a communicative framework. Lexicographers have written a number of texts collected in individual dictionaries for the purpose of helping users by sharing knowledge. Bergenholtz, Tarp and Wiegand (1999: 1763) suggest that a dictionary is not a text in itself but

rather a collection of text types or text genres with different communicative functions. It can be argued that the communication process starts when users formulate questions to which they seek answers in their dictionaries and ends when they have found the answers or not. However, it is important to note that users often see the texts in a dictionary as typical examples of a one-way communication: lexicographers communicate data to dictionary users but do not engage in an instant dialogue, nor intend to do so. This is wholly true of printed dictionaries where no one expects lexicographers and users to enter into dialogues during ongoing acts of consultation. Some dictionaries invite users to contact lexicographers or publishers if they have any comments, but this does not form part of specific consultative acts in which users seek immediate help.

It is debatable whether electronic dictionaries allow users to enter into dialogues with lexicographers. Many intranet and internet dictionaries allow users to contact their lexicographers, but whether they receive a reply, and, if so, when, is uncertain. At present, it is fair to say that no electronic dictionaries allow users to enter into direct dialogues with lexicographers in specific instances of consultation so that, during an ongoing act of consultation, users can ask clarifying questions and receive answers immediately. If dialogues are established they will occur after the consultative acts have ended. The future development of interactive online systems may change the dictionary environment, but until then, users are unlikely to regard dictionary consultation as a two-way communication. Lexicographers should take this restricted communicative framework into consideration when planning and compiling dictionaries in order to help users.

3.   What are lexicographical information costs?

The costs of acquiring information are generally a natural part of the process and therefore we do not notice them. However, if they do not form part of the natural process they interrupt it, thereby becoming salient, causing users to notice them. In order to establish a workable basis for analyzing the ease or difficulty facing users of dictionaries in their attempts to find help, it is necessary to determine the types of costs involved. Economists were some of the first to use and analyze the concept of information costs, applying it in business economics and in management theory. Example 1 contains a general definition:

> **information cost**
> The cost to an organisation of obtaining
> knowledge of its business environment.

Example 1: Article from *Routledge Dictionary of Economics* (2002).

At first glance, the emphasis in Example 1 appears to be on costs measured in monetary terms, but the concept is somewhat wider. As pointed out by Casson

(2000: 117-122), information costs include what may be called *temporal costs*, i.e. the time spent searching, looking up and understanding data in addition to *pecuniary costs* incurred in appraising and planning activities, exchanging information, and verifying data. Even though this perspective has an economic bias, information costs are also relevant in non-economic settings.

The concept of information costs has come to be used outside economic and management contexts, for instance in information theory, which is reflected in the following definition:

information costs
> The costs of acquiring, processing, and using information, part of the costs associated with any type of decision, including a decision how to vote.

Example 2: Article from *Texas Politics* (2007).

Example 2 contains two important points: (a) information costs are part of acquiring information; and (b) information costs are part of processing the relevant data. The definition introduces a third element, which may be referred to as *intellectual costs*, i.e. the mental efforts involved in processing the data. Applied to dictionary consultation, this means that it involves costs to obtain the assistance sought from dictionaries, and to decode the data for finding a useful interpretation.

Lexicographers should carefully consider the ease with which users will be able to find help from their dictionaries. One way of dealing with this issue is to attempt to adapt the concept of information costs to lexicography because dictionaries are carriers of texts. For the purposes of dictionary compiling, use, analysis and research, Nielsen (1999: 111) defines *lexicographical information costs* as the effort that a user believes or feels is associated with consulting a dictionary, an article or any other text part of a dictionary. For the user the important point is the relation between the anticipated information costs and the anticipated information value, i.e. the user's gain from consulting a dictionary in a given consultative act.

Lexicographical information costs may be divided into two distinct types. First, *search-related information costs* are the costs (i.e. effort) related to the look-up acts users have to perform when consulting a dictionary in order to gain access to the data for which they are searching. The more acts the user has to perform to reach the target, the more significant will be the costs, and so will be the risk of going wrong at one or more steps in the search process. The very design and structure of the dictionary may contribute to keeping lexicographical information costs at a low level, as an inappropriate design and structure may lead to high or increasing information costs. De Schryver (2003: 173-182) provides a good account of search options in electronic dictionaries, and typical examples of situations involving search-related information costs comprise: cross-references included without any clear purpose; the number of individual

steps necessary in the look-up process (in printed and electronic dictionaries) before the sought goal is reached; the need for scrolling the screen to find a specific item; and the possibility of searching for a specific word instead of scrolling a long list.

The search options available in electronic dictionaries directly affect the users' perception of costs associated with finding help to solve their problems. It is important to users whether they can only search for full, lemmatized terms, whether they can also search for parts of lemmatized terms (e.g. 'begins with', 'ends with' and 'includes'), or whether they can search for full and partial terms in any part of dictionary articles (e.g. in definitions and in collocations) as well as in the outer texts, such as subject-field components, style guides and different kinds of tables. The above examples may lead some to conclude that all lexicographical information costs are inherently quantitative and temporal in nature, but this is not the case.

The second type of lexicographical information costs may be described as having a qualitative nature. *Comprehension-related information costs* are the costs (i.e. effort) related to the user's ability to understand and interpret the data presented in a dictionary. In other words: How easy or difficult is it for users to process the data in the dictionary? This is directly related to the factual and linguistic competence of users and the way in which the data are presented. The wording and presentation of data in the articles, for instance a high degree of textual condensation in definitions, may increase the information costs. It is clear from the above description of lexicographical information costs that the concept relates to the search for and interpretation of texts or text parts in dictionaries, but it is relevant to determine its sphere of application.

The scope of application is generally the dictionary as a whole, i.e. the entire collection of text types or text genres. However, lexicographical information costs do not directly apply to all of these text components. Tarp (1998: 122-131) distinguishes between *lexicographical components*, which are related to the use and functions of dictionaries, and *extra-lexicographical components*, which are not related to the use and functions of dictionaries. Examples of components irrelevant to the use and functions of dictionaries are acknowledgements, picture credits, eulogistic forewords (such as that found in the *Oxford Advanced Learner's Dictionary*), and single user licence agreements for accompanying CD-ROMs (such as that found in the *Longman Dictionary of Contemporary English*). In the context of lexicographical information costs, only lexicographical components warrant attention, the following being some of the most important:

— Front matter: user guide, subject-field component, writing guide

— Wordlist: articles, inserts

— Back matter: appendices

It follows from this that lexicographers should see their dictionaries as collections of interrelated texts. The above list uses the traditional terminology appli-

cable to printed dictionaries, but this does not exclude electronic dictionaries. On the contrary, authors and users of CD-ROM, intranet and internet dictionaries are also affected by lexicographical information costs as these types of reference works normally each contains several lexicographical components, though their actual positions in relation to the wordlist are difficult to determine. As argued in Nielsen (2006: 10-18), this means that outer texts such as the user guide should be easy to read and actually show how users may find the data presented in the dictionary. For instance, a user guide containing a balanced combination of visual keys and supporting text is usually easier to read and understand than one written in a condensed, straight-text mode. Similarly, the user guide should explain why the different lexicographical components are included, how they supplement each other, and how the components support the functions of the dictionary.

The discussion so far indicates that the use of dictionaries is not cost free and that it is impossible to eliminate information costs. These costs, however, can be minimized. Furthermore, dictionaries are not merely wordlists with adjunct texts but collections of interrelated texts supporting use and functions. Finally, it should be appreciated that a user's choice of dictionary depends on the relation between the expected information costs and the expected informative value when consulting a dictionary. Every time we consult dictionaries for help, we usually have a favourite dictionary, which we consult before any other. The question then is: what is actually in the dictionary?

4.   An important distinction between data and information

In our daily lives, we normally use the two words *data* and *information* interchangeably. Nevertheless, when we examine, describe, use and analyze dictionaries in an academic or a professional context it is important to distinguish between the two. As argued by Wiegand (2000a: 22-25; 2005: 216), a dictionary does not contain information, but data that users can convert into information through a mental process and thereby satisfy their needs for help in a given situation. This distinction is also relevant to the discussion of lexicographical information costs, because it sheds light on the way in which users read and understand what is available in dictionaries. Many people have attempted to define data, Example 3 providing a generally accepted definition.

data
        any representation such as characters or analog quantities
        to which meaning is, or might be, assigned.

Example 3: Article excerpt from *Dictionary of Banking and Financial Services* (1985).

This definition indicates that data items do not carry information, but that meaning may be assigned to them. According to Liu (2000: 1), 'information is

processed data that has meaning to its users' and 'all information is "carried" by signs of one kind or another'. By analogy, this means that lexicographical data may be treated as semiotic signs carrying information. When this is assumed, several theoretical approaches are available. It is difficult to say which is the right approach, but Peirce (1992) offers one that specifically deals with the process of assigning meaning to signs through the interplay between what is referred to as universal categories. Peirce (1992: 296) describes the three universal categories called firstness, secondness and thirdness as follows:

> First is the conception of being or existing independent of anything else. Second is the conception of being relative to, the conception of reaction with, something else. Third is the conception of mediation, whereby a first and second are brought into relation.

Merrel (2000: 32-33) applies the universal categories to various genres, including poetry and art, on the premise that firstness is possibility, secondness is actuality, and thirdness is potentiality, probability or necessity. Based on his application of Peirce's universal categories to texts, the following is an attempt to describe how meaning is assigned to lexicographical data:

(a)   A sign is a representation on paper/screen and has the possibility of meaning something to someone.

(b)   The actual reading of the sign by the dictionary user creates one or more relations between the sign and the user/reader triggering something in the user's mind.

(c)   The user's reading of the sign and his mindset, his experience and conventions combine to establish a meaning of the lexicographical sign.

This description of data as signs involves two important aspects. Firstly, it can help to understand why different users may interpret the same text or text part differently, depending on their mindsets, experience and the conventions they know and apply. Secondly, data items expressed as signs may take different forms, for instance numerical figures, alphabetical letters, strings of graphemes, graphical representations, symbols, icons and textual suprasegments. I suggest that lexicographers should take both aspects into consideration when writing their dictionary texts with a view to easy consultation. By their very nature, lexicographical information costs cannot be eliminated, but it is appropriate to look at some factors that may help to keep them at a reasonable level.

5.   Article structures may reduce information costs

The access route to the data in dictionaries may reduce or increase lexicographical information costs. As shown in section 3, the wordlist and outer texts fall within the scope of information costs and the individual lexicographical

components belong to different text genres. This means that the access structure applying to one component may differ from that applying to another. The central component of a printed dictionary is the wordlist and the end result of a successful search is a text, usually a dictionary article. Most CD-ROM and many online dictionaries are exclusively or substantially based on printed dictionaries whose data have been transferred to electronic media. Consequently, the outcome of a successful search in these dictionaries is usually also a text in the form of an article. Advanced online dictionaries may not have wordlists but databases in which the search is conducted electronically and the result of a successful search may be a screen representation of data, including photographs and video footage, collected from various places in the databases. Nevertheless, it seems reasonable to say that many search results are presented as text in online dictionaries, though perhaps not as dictionary articles proper.

Users place easy and quick access to data items in articles high on their list of priorities. By using standardized article structures in dictionaries, lexicographers make it easy for users to familiarize themselves with the layout of articles. They may divide the text of an article into fields, or zones, that each contains a specific type of data. Users can easily navigate their way through such standardized articles because it involves less effort or incurs a lower cost to find what they are looking for, or to realize that what they are looking for is not in the article.

Each article in a dictionary is different from all other articles, but they may be analyzed using the same approach to access. Wiegand (2000b: 259 et seq.) suggests that articles may be analyzed in terms of search zones, which are relevant when considering search-related information costs. This may be illustrated on the basis of the following article:

final dividend
⟨a, the, -s⟩

endeligt udbytte
⟨et; det endelige udbytte, endelige udbytter, de endelige udbytter⟩

Definition
Final dividend is remaining dividend payable at the end of the financial year. The final dividend constitutes the difference between any interim dividend paid during the year and the total dividend declared at a company's annual general meeting.

Collocations
• declare final dividend of 10%
deklarere endeligt udbytte på 10%
• recommend a final dividend of 9.5%
foreslå et endeligt udbytte på 9,5%
• entitled to the 2006 final dividend
berettiget til endelig udbytte for 2006
• proposed final dividend
forslag til endeligt udbytte

See also:
interim dividend

Example 4:   Article from *English–Danish Accounting Dictionary* (2005-2008).

This article comes from an online English–Danish accounting dictionary, hence the somewhat expanded layout. A closer study shows that it has been divided into separate search zones based on the types of data they contain. The data should be presented as separate search zones within which the user can easily look for individual data items. The way to create search zones in a dictionary article is to make sure that the article text does not appear as a long, unbroken chain of words, signs, etc. This may be achieved by using various structural indicators as signposts such as blank space at the end of a line, typographical symbols, icons, italics and boldface. One result is that users may save time by proceeding directly to the relevant search zone without having to go through the article from the beginning until they reach the zone containing the data they are looking for.

The search zone structure in Example 4 is vertical because all search zones are placed so that they are either above or below one or more other search zones. The first search zone contains the lemma and the item showing the inflectional paradigm; the suprasegmental feature boldface is used as a search zone indicator. Search zone two is made up of the Danish equivalent and the item indicating the inflectional paradigm; boldface is again used as a search zone indicator. The third search zone contains the English definition and is introduced by the headline 'Definition' in a different colour (in this case blue). Search zone four presents the English collocations and phrases and their Danish translations. Again a headline 'Collocations' in a different colour is used as search zone indicator. The fifth search zone has a cross-reference to another lemma, the headline 'See also' in blue functioning as search zone indicator. Users can click on the cross-reference address and be linked directly to the article **interim dividend**. The blank space after the last line of the zone and a blank line between zones mark the end of each search zone.

Some of the search zones in Example 4 are divided into subzones that may facilitate access to the data and reduce temporal costs. For example, the first search zone is divided into two subzones, the first consisting of the boldface lemma followed by the grammatical data within angle brackets. Search zone four contains collocations and phrases in four separate subzones, each introduced by a bullet and with the English collocations against a grey background as signpost. Finally, the overall search path users have to follow consists of the search zone indicators, is positioned vertically and is margined to the left of the article. If the text had been written as a long, unbroken chain of words, signs, etc., the search path would have been sinuous. Even though the article in Example 4 comes from an online dictionary, this description shows that it is possible to analyze the elements of articles in printed and electronic dictionaries in terms of textual topology, search zones and subzones and to arrange

these elements in such a way as to ensure the continuity of the article text and easy access to the data.

Search-related information costs are related to the lexicographical signs introduced in section 4. Signs may appear in various shapes and forms but their meanings depend on the knowledge and experience of readers as well as the conventions they know and apply. For instance, the use of bullets to signpost the beginning of a collocation or phrase is a convention established by the lexicographers of the *English–Danish Accounting Dictionary* and has to be learnt by users of this particular dictionary. Other lexicographers may employ other signposts in their dictionaries. A blank at the end of search zones is a more widely used sign in dictionary articles and may be said to be a generally applicable convention. The use of a suprasegment, for instance boldface, to signpost the lemma is perhaps a universal convention and known by all users of dictionaries. Users generally widen their experience as they consult different dictionaries and thereby familiarize themselves with common and universal signs and conventions, whereas dictionary-specific signs and conventions will have to be learnt on a case-by-case basis. By properly using such signposts in articles, lexicographers can keep the search-related information costs at a low and reasonable level.

6.   Dictionary functions and use-situations as benchmarks

Properly arranged textual structures may well facilitate the access to lexicographical data but they do not guarantee the easy understanding of data. The study of comprehension-related information costs generally involves what may be called a qualitative approach. Users reading the data presented in dictionaries need to process these data into information useful to them. The question is what factors to take into account when studying the various reasons causing the information generating process to be more or less costly. One way in which to analyze and evaluate information costs is to relate the data to the functions of dictionaries. In order to do this, it will be necessary to take a closer look at the functions of dictionaries and the needs of users. The function of dictionaries is to provide assistance to particular user groups with specific characteristics in order to cover the needs that arise in specific types of use-situations. The types of use-situations allow lexicographers to distinguish between two overall functions.

Bergenholtz and Nielsen (2006: 286) suggest that dictionaries may be designed to deal with two general types of functions. Firstly, *communication-orientated functions* are those that assist users in solving problems related to the understanding of texts, the production of texts and the translation of texts. For example, in an ongoing written or spoken act of communication between two or more persons a communication-related problem arises and the interlocutors consult a dictionary in order to find help. Secondly, *cognition-orientated functions* are those that provide general and encyclopedic information, special informa-

tion about a topic, or information about language in general or in specific contexts. Unrelated to any ongoing communication situation, a person may want to obtain additional information on a topic such as general cultural information or specialized information concerning an issue within a particular subject-field. In order to find help in such situations, we often consult dictionaries because we think that they can help us solve our problems.

In order to meet the needs of users in specific types of use-situations, lexicographers should consider profiling user groups. Bergenholtz and Nielsen (2006: 285) provide the following list of characteristics lexicographers may take into consideration in an attempt to match user needs and dictionary functions:

— Which language is their native language?
— At what level do they master their native language?
— At what level do they master a foreign language?
— How extensive is their experience in translating between the languages in question?
— What is the level of their general cultural and encyclopedic knowledge?
— At what level do they master the special subject field in question?
— At what level do they master the corresponding LSP in their native language?
— At what level do they master the corresponding LSP in the foreign language?

There may be other relevant types of characteristics for a particular dictionary which do not all apply to all dictionaries, but those listed above are the most important for profiling a specific user group. The important point is that the relation between dictionary functions, user needs and use-situations may serve as benchmarks for assessing comprehension-related information costs, because they can act as standards against which to compare the costs involved.

The real needs of users directly affect the practical work of lexicographers. Laypersons using specialized dictionaries have characteristics, i.e. factual and linguistic competences, differing from those of experts using the same or similar dictionaries. One way of dealing with this is to present the data in ways adapted to the intended user groups. As indicated in section 3, this applies to all lexicographical components. Bergenholtz and Kaufmann (1997: 115-116) illustrate how definitions may be adapted to the user profile of an electronic dictionary of gene technology.

One advantage of electronic dictionaries is that they may allow users to select particular functions for specific groups of users. A layperson having difficulties understanding a newspaper article or a popularized book on gene technology may seek help in a dictionary. He/she needs to know what a gene is. Having found an electronic dictionary designed to help laypersons understand texts, he/she searches for the word *gene* and finds the following result:

gene
>   the basic unit of inheritance which is transmitted from
>   parents to offspring.

Example 5: Imaginary dictionary article helping laypersons understand texts on gene technology.

The article in Example 5 presents the data in a way that is as easy as possible for laypersons to understand. It is factually correct, devoid of technical jargon and should help users to follow texts in which the term *gene* occurs. In contrast, user profiles of experts and semi-experts show that these user groups have factual and linguistic competences different from those of laypersons and therefore have different needs. A semi-expert may have difficulty understanding a highly scientific text on gene technology and wants to know what a gene actually is. He/she consults the same electronic dictionary and selects a set-up for semi-experts wanting help to understand texts. Example 6 shows the result of his/her search.

gene
>   A gene is a DNA sequence ending an mRNA (protein),
>   tRNA or rRNA.

Example 6: Imaginary dictionary article helping semi-experts understand texts on gene technology.

The definition in Example 6 is quite different from that written for laypersons. It uses the correct scientific terms, which makes it impossible for laypersons to understand. The comprehension-related information costs for laypersons are prohibitive, whereas they are reasonable for experts and semi-experts. The factual and linguistic competences of intended user groups and their needs in different use-situations directly affect dictionary functions and the ways in which lexicographical data should be presented. Moreover, a reasonable level of comprehension-related information costs does not hinge on definitions being short or long; the definitions belong to different text genres. The relationship between dictionary functions, user needs and use-situations may therefore be used as benchmarks in assessing comprehension-related information costs.

7.   Textual condensation may increase information costs

The condensation of lexicographical texts or text parts directly impacts on the process that converts data into useful information. The immediate effect of textual condensation is the reduction of the length and the number of characters of natural sentences or sentence parts for the purpose of saving space. A typical example is the use of the tilde, or swung dash, to replace the lemma inside the article text, usually in collocations and phrases. Lexicographers have a range of options to choose from when condensing texts, Wiegand (1996: 139)

offering the following list: shortening, abbreviating, omitting, shifting, substituting, summarizing and embedding. These condensing operations all involve different types of textual reduction and may be used in printed as well as electronic dictionaries; they are generally found in dictionary articles.

Seen from the user's point of view, condensing operations carried out in dictionary articles may be more or less successful. Some condensed texts are easy to decode, whereas others are difficult to understand. In any case, the reduction of natural sentences to shorter, non-natural sentences or fragments affects comprehension-related information costs, as illustrated in Example 7:

> **hamstern** [ˈhamstɐn], ‹reg. Vb.; hat› umg. emot. neg.
>   /jmd./ *etw.* ~ SYN ˈetw. hortenˈ; ↗ FELD I.16.2:
>   *Zucker, Mehl, Lebensmittel* ~ ❖ ↗ **Hamster**

Example 7: Article from *De Gruyter Wörterbuch Deutsch als Fremdsprache* (2000).

The article in Example 7 clearly shows that the textual condensation carried out has resulted in non-natural text. The lexicographers may very well feel that their condensing operations (omission, abbreviating, and substituting) have been successful, but most users will probably disagree. The dictionary is intended for text production by learners of German as a foreign language and it seems reasonable to conclude that the comprehension-related information costs are high. In order to help users whose native language is not German, the dictionary contains a list of abbreviations used, but to be able to read the articles fairly easily, these non-native speakers have to learn and memorize no less than 144 abbreviations and their meanings. Even native speakers of German may find this arduous work. In dictionaries designed for the productive needs of non-native speakers of a particular language, non-natural texts or text parts in articles should be used sparingly and with due regard to the factual and linguistic competences identified in the user profile.

Dictionaries designed to provide assistance in communication-orientated use-situations such as text production often contain competence or editorial examples. As illustrated in Example 8, these examples show users how words can be used in connection with other words.

> **Wörterbuch** […]
>   *ein medizinisches, englisches* ~ *; ein historisches,*
>   *ein-, zweisprachiges* ~ *; […]*

Example 8: Article excerpt from *De Gruyter Wörterbuch Deutsch als Fremdsprache* (2000).

Example 8 contains a typical instance of condensed items giving examples of use; the condensing operations carried out are omitting and substituting. The textual condensation has resulted in non-natural partial sentences, e.g. *ein medizinisches, englisches* ~ instead of *ein medizinisches Wörterbuch oder ein englisches Wörterbuch*. And can non-native speakers of German really be blamed if,

after having consulted the above article, they produce words such as *ein-Wörterbuch* and *ein ein-Wörterbuch*? As the presentation of such examples in production dictionaries intended for learners of foreign languages burdens users with high information costs, lexicographers should not condense competence examples in this way.

There is no established and generally accepted yardstick for measuring textual condensation in dictionary articles. Nielsen (2002: 606) proposes the use of a *textual condensation ratio* as an indication of the degree of textual condensation carried out. The textual condensation ratio is calculated as the number of condensing operations divided by the number of elementary items in an article. Condensation ratios below 1 indicate that the number of condensing operations performed is smaller than the number of elementary items. In other words, not all elementary items have been condensed. In contrast, condensation ratios above 1 indicate that the number of condensing operations is larger than the number of elementary items. This means that at least one item has been subject to two or more condensing operations. Although this is a rather crude measure, it points to the level of textual condensation in dictionary articles and moreover also indicates whether this has a significant or insignificant effect on comprehension-related information costs.

8.   Cross-references and data distribution affect dictionary use

Lexicographical structures may be used to marshal a dictionary's data into fulfilling its function or functions. Two prominent structures are the distribution structure and the cross-reference structure. The former decides where in the dictionary data is placed, for instance some data about a term may be placed in an article and some in an outer text. The cross-reference structure, also known as the mediostructure, is a network structure that consists of the set of relations existing between data by way of cross-references. An analysis of these two structures involves access to and understanding of data, which in turn affect information costs.

Cross-references are widely used in dictionaries to help users acquire further information. It is possible to distinguish between different types of cross-references: cross-references within articles, cross-references from one article to another, cross-references from articles to outer texts or vice versa, cross-references within outer texts, cross-references from one outer text to another, and cross-references to places outside the dictionary. Nielsen (1999: 98-99) shows that it may be beneficial to relate these types of cross-references to dictionary functions, summarized as follows:

—   cross-references supporting comprehension
—   cross-references supporting text production
—   cross-references supporting translation
—   cross-references supporting knowledge acquisition

Cross-references should not send users on random searches through the word-list or outer texts. Information overload caused by never-ending cross-reference chains may cause users to end up perplexed and unsure of why they have been referred to different places in the dictionary and what the relationships are between the terms and data concerned. In such cases, temporal costs are involved and so are intellectual costs. One way to minimize lexicographical information costs is for lexicographers to use focused cross-references which support the principal function or functions of the dictionary in question. For example, dictionaries designed for translating may include function-related cross-references enabling users to establish links between data supporting the translation of texts. Lexicographers of polyfunctional dictionaries may elect to use cross-references supporting the primary function only.

Cross-references may be used to establish various types of relationships depending on the functions of dictionaries. One type is hierarchical relationships between lemmatized terms, which may help users understand texts or gain knowledge. It is important to remember that cross-references do not themselves create hierarchical relationships, but connect data in different places enabling users to recreate the relationships on the basis of the data and the links. Example 9 contains three articles illustrating this principle.

> **defendant** *n* a party against whom a criminal charge
>   or civil claim is made: compare PLAINTIFF.
>
> **litigant** *n* a person engaged in a lawsuit.
>
> **plaintiff** *n* somebody who brings a civil legal action
>   against another in a court of law: compare
>   DEFENDANT.

Example 9: Articles from *Penguin Office Dictionary* (2006).

Two of the articles in Example 9 contain cross-references to each other and help users to establish a relationship between the terms *defendant* and *plaintiff* from the data in the articles. However, the full potential of using cross-references to assist users in understanding conceptual relationships has not been realized. If the article **litigant** had contained cross-references to the other two articles and these in turn had referred to **litigant**, users would have been able to establish the hierarchical relationship between the three terms. The relationship represented by the data in the referential network would show a species–genus relation, with *litigant* as the generic term and *defendant* and *plaintiff* as specific terms (in this case co-hyponyms). This would have saved users the extra time and effort involved before deducing this terminological hierarchy on the basis of other data that would have to be gathered from elsewhere.

Hierarchical relationships may also be established in other ways depending on the distribution structure and dictionary functions. The data in Example 9 were distributed so that the three articles contain different explanatory data

with 'horizontal' cross-references between the specific terms. Another distribution structure may cause the data to be placed differently:

defendant *n* see LITIGANT.

litigant *n* A litigant is a person engaged in a lawsuit.
> A civil lawsuit involves two parties, a plaintiff, who brings a civil legal action against another party called a defendant. In a criminal case the prosecution brings a criminal charge against a party called a defendant.

plaintiff *n* see LITIGANT.

Example 10: Alternative arrangement of data in the articles shown in Example 9.

In the articles in Example 10, the distribution structure has resulted in two cross-reference articles and one synopsis article. Dictionaries designed for helping users to understand texts and develop knowledge may benefit from this type of arrangement where all explanatory data have been placed in one article, and where the cross-reference structure is 'vertical' or 'bottom-up', i.e. from the specific terms to the generic term. Users should be able to create the terminological hierarchy between the three terms without spending too much time and effort. Finally, signposts such as suprasegmental features, for instance italics or colour, could have been used to reduce the time spent on finding the terms *plaintiff* and *defendant* in the synopsis article.

Cross-references may be used to establish links between data in articles and in outer texts. The distribution structure may cause some of the data to be placed in, for instance, a subject-field component and not in the articles. Dictionaries designed for understanding and knowledge building will then need cross-references linking the data so that users may establish the relevant relationship. In some cases, explanatory data in subject-field components may supplement data in articles, and in other cases articles contain no explanatory data; the explanatory data are located in the subject-field component giving an introduction to or an overview of the subject-field treated by the dictionary. Example 11 contains articles and a relevant excerpt from a subject-field component illustrating the interrelationship of the distribution and linking of data.

Civil Division *s* → §2
Court of Appeal *s* → §2; §14
Lord Justice of Appeal *s* → §2; §14
Master of the Rolls *s* → §2

§2 Court of Appeal, Civil Division
The civil division of the appeal court ranks second in the English court hierarchy. Court of Appeal, Civil Division has only jurisdiction to hear civil appeal cases from England and Wales appealed from the High Court and county courts. The president of the court is the **Master**

> **of the Rolls**, and the ordinary appeal judges are called **Lords Justices of Appeal**. Appeal cases are normally heard by 3 judges and appeal lies to the House of Lords if permission to appeal is granted by either the Court of Appeal or the House of Lords.

Example 11:   Articles and excerpt from subject-field component in *Engelsk–Dansk Juridisk Basisordbog* (1993).

The four articles in Example 11 are shown as they appear in the dictionary. They do not contain any explanatory, factual data but only cross-references to numbered sections in the subject-field component where the necessary data are given. (Note that this dictionary was designed for Danish users and the subject-field component is written in Danish but has, for practical reasons, been translated into English.) Section 2 of the subject-field component explains the term *Court of Appeal, Civil Division* and the relevant judges. Accordingly, the data fulfil a cognition-orientated function. Moreover, the original Danish text suggests possible translations of the boldface terms, thereby fulfilling a communication-orientated function (*the president* and *ordinary appeal judges*).

Cross-references that support understanding and knowledge development need to work optimally, and lexicographers should carefully consider the search process involved. The cross-reference article **Civil Division** directly specifies its destination, i.e. a numbered section in the subject-field component. The purpose of the article is to give information about the lemma, referring users to a place in an outside text where the relevant data are located. As *§2* in the subject-field component deals with the concept of the *Court of Appeal, Civil Division* in an overall way, this type of cross-reference may be called direct. The last two articles in Example 11 contain indirect cross-references because their addresses indicate the location of an overall text part within which the actual addresses are located. The purpose of the article **Master of the Rolls** is to explain the lemma to users by referring them to a place in the outer text. The direct cross-reference address is *§2*, and the indirect address of *Master of the Rolls* is the identical linguistic unit written in boldface in that section. In order to facilitate access the subject-field component is divided into numbered sections with headlines in boldface and all addresses inside these sections are also written in boldface. This reduces the time spent by users searching for sections and terms explained within sections.

Hierarchical relationships do not only involve genus–species relations. Cross-references may also be used to establish links enabling users to create part–whole relations and generic topic–proper name relations. Moreover, cross-references can be used to link data describing non-hierarchical relationships such as sequential relations. Schaeder (1995: 131-132) shows that two types of sequential relationships are relevant: those based on the linear progression of actions, events, etc. in time or space, and those based on thematic connections (see also Nielsen 1999: 104-105). No matter which function cross-references and the distribution of data support, it is imperative that lexicogra-

phers take care to provide the data in the articles or outer texts necessary for users to establish the hierarchical or sequential relationships. The underlying principle should be that users experience a positive informative value every time explicit cross-references are followed so that any information costs are part of the natural flow of the process.

9.   The dual function of lexicographical information costs

The concept of lexicographical information costs is applicable in a range of situations. Its main operative feature is that it has dual applicability in a temporal sense: it can be applied to the time before and the time after the completion of dictionaries. *Retrospective application* is mainly descriptive because it refers to situations that occur after the completion and publication of dictionaries and relates to lexicographical tools already in existence. Examples of retrospective application are: dictionary reviews, metalexicographical analyses and academic descriptions of dictionaries. Reviewers may assess dictionaries on the basis of search-related and comprehension-related information costs in addition to other relevant criteria and issues; and metalexicographers may analyze existing dictionaries and draw attention to positive and negative cost effects.

    Prospective application is generally generative as it refers to situations before the completion and publication of dictionaries and relates to the production of future lexicographical tools. Examples of prospective application are: the planning phase and the compilation phase of dictionaries. Practical lexicographers may prepare instructions for the dictionary work designed to keep information costs at a reasonable level in respect of all the intended dictionary functions and user groups. Theoretical lexicographers may develop new principles for dictionary compiling, making it possible to reduce function-related information costs relative to existing levels and propose new standards for measuring information costs. To sum up, the concept of lexicographical information costs applies to diverse situations that permeate the world of dictionaries, whether such costs go unnoticed or disrupt the natural flow of dictionary consultation, analysis, research or assessment.

10.   Conclusion

When we analyze, compile, research or use dictionaries, lexicographical information costs are implicitly or explicitly part of what we do. Users of dictionaries often choose their favourite among their dictionaries when they need help to solve specific problems in specific types of use-situations. Their choice normally depends on the ease and convenience they know or believe are associated with a particular dictionary. Users within the relevant user group elect to consult *Encyclopedia Britannica 2007 Children's Encyclopedia* instead of *Encyclopedia Britannica*, because they know or believe that they can find answers to

their questions without incurring unreasonably high information costs. Lexicographers should strive to present the data in their dictionaries in such a way that users feel they find answers to their questions easily and gain useful knowledge by consulting the dictionaries. Owing to their very nature lexicographical information costs cannot be eliminated, but it is appropriate to draw attention to the fact that prudent and proper consideration may result in a reasonable cost level not seriously affecting the use of dictionaries.

In assessing lexicographical information costs, several factors come into play. Whether costs are high or low in a given situation depends on several interrelated factors, such as the user group, the factual and linguistic competences of users, the structure of data in articles and outer texts, the time it takes to locate the data, the degree of the condensation of data, the distribution of data and the links connecting them. A number of yardsticks have been proposed for measuring search-related and comprehension-related information costs in an attempt to answer the question: how easy (or difficult) is it for users to find and understand the data presented? It has thus been possible to lay the foundation for a framework for evaluating lexicographical information costs for the purpose of dictionary making, use, analysis and research.

References

Dictionaries

De Gruyter Wörterbuch Deutsch als Fremdsprache 2000 = **Kempcke, Günther (Ed.).** 2000. Berlin/New York: De Gruyter.

Dictionary of Banking and Financial Services 1985 = **Rosenberg, Jerry M.** 1985. New York: John Wiley & Sons.

Encyclopedia Britannica 2007 = **Encyclopedia Britannica, Inc.** 2007.

Encyclopedia Britannica 2007 Children's Encyclopedia 2007 = **Encyclopedia Britannica, Inc.** 2007.

Engelsk–Dansk Juridisk Basisordbog 1993 = **Nielsen, Sandro.** 1993. Copenhagen: Munksgaard.

English–Danish Accounting Dictionary 2005–2008 = **Nielsen, Sandro, Lise Mourier and Henning Bergenholtz.** 2005–2008. http://www.regnskabsordbogen.dk/regn/gbgb/gbregn.aspx [25 August 2008].

Oxford Advanced Learner's Dictionary 2005 = **Wehmeier, Sally (Ed.).** 2005. Oxford: Oxford University Press.

Penguin Office Dictionary 2006 = **Allen, Robert (Ed.).** 2006. London: Penguin Books.

Routledge Dictionary of Economics 2002 = **Rutherford, Donald.** 2002. London/New York: Routledge.

Other literature

Bergenholtz, Henning and Uwe Kaufmann. 1997. Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. *Hermes* 18: 91-125.

Bergenholtz, Henning and Sandro Nielsen. 2006. Subject-field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12(2): 281-303.

Bergenholtz, Henning, Sven Tarp and Herbert Ernst Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, Lothar, Hartwig Kalverkämper, Herbert E. Wiegand together with Christian Galinski and Werner Hüllen (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research, Bd./Vol. 2*: 1762-1832. Berlin/New York: De Gruyter.

Casson, Mark. 2000. *Economics of International Business. A New Research Agenda*. Cheltenham: Edward Elgar.

De Schryver, Gilles-Maurice. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

Liu, Kecheng. 2000. *Semiotics in Information Systems Engineering*. Port Chester, NY: Cambridge University Press.

Merrel, Floyd. 2000. Charles Sanders Peirce's Concept of the Sign. Copley, Paul (Ed.). 2000. *Routledge Critical Dictionary of Semiotics and Linguistics*: 28-39. Florence, KY: Routledge.

Nielsen, Sandro. 1999. Mediostructures in Bilingual LSP Dictionaries. *Lexicographica. International Annual for Lexicography* 15: 90-113.

Nielsen, Sandro. 2002. Textual Condensation in the Articles of *de Gruyter Wörterbuch Deutsch als Fremdsprache*. Wiegand, Herbert E. (Ed.). 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des »de Gruyter Wörterbuchs Deutsch als Fremdsprache«*: 597-608. Tübingen: Max Niemeyer.

Nielsen, Sandro. 2006. A Functional Approach to User Guides. *Dictionaries. Journal of the Dictionary Society of North America* 27: 1-20.

Peirce, Charles Sanders. 1992. The Architecture of Theories. Houser, N. and C. Kloesel (Eds.). 1992. *The Essential Peirce. Selected Philosophical Writings, Vol. 1*: 285-297. Bloomington/Indianapolis: Indiana University Press. (First published in *The Monist* 1, 1891: 161-176.)

Schaeder, Burkhard. 1995. Mediostrukturen in Fachwörterbüchern. *Lexicographica. International Annual for Lexicography* 11: 121-134.

Tarp, Sven. 1998. Leksikografien på egne ben. Fordelingsstrukturer og byggedele i et brugerorienteret perspektiv. *Hermes* 21: 212-137.

Texas Politics. 2007. *Key Words and Phrases*. http://texaspolitics.laits.utexas.edu/6_keywords.html [25 August 2008].

Wiegand, Herbert Ernst. 1996. Textual Condensation in Printed Dictionaries. A Theoretical Draft. *Lexikos* 6: 133-158.

Wiegand, Herbert Ernst. 2000a. Wissen, Wissensrepräsentation und Printwörterbücher. Heid, Ulrich et al. (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8–12, 2000*: 15-38. Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.

Wiegand, Herbert Ernst. 2000b. Über Suchbereiche, Suchzonen und ihre textuellen Strukturen in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. Wiegand, H.E. (Ed.). 2000. *Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 233-301. Tübingen: Max Niemeyer.

Wiegand, Herbert Ernst. 2005. Angaben, funktionale Angabezusätze, Angabetexte, Angabestrukturen, Strukturanzeiger, Kommentare und mehr. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica. International Annual for Lexicography* 21: 202-344.