Making a dictionary app from a lexical database: the case of the Contemporary Dictionary of the Swedish Academy

Louise Holmer, Monica von Martens, Emma Sköldberg

Department of Swedish, University of Gothenburg PO Box 200, SE-405 30 Gothenburg, Sweden E-mail: louise.holmer@svenska.gu.se, monica.martens@gu.se, emma.skoeldberg@svenska.gu.se

Abstract

Developing an app version of a printed dictionary is a new challenge faced by lexicographers. Lexicographers involved in the app development process must consider fundamental lexicographic aspects as well as learn to understand technological and usage issues inherent to the new media. An inevitable question is how closely the content and layout can be made to match the printed dictionary while still offering 'digital' functionality such as linking, collapsed sections, audio, etc. Only a few reports discussing these issues have so far been published.

The aim of our paper is to further advance the exchange of knowledge and experience by sharing our observations made during the development of a new app corresponding to the comprehensive printed dictionary, *Svensk ordbok utgiven av Svenska Akademien* (the Contemporary Dictionary of the Swedish Academy, 2009). The app is the result of close cooperation between the financer The Swedish Academy, lexicographers and system developers at the Department of Swedish, University of Gothenburg and Isolve AB, a Stockholm-based app development agency specializing in dictionary apps.

Keywords: dictionary app; electronic dictionary; dictionary application; lexical database

1. Introduction

As the extent of use of smartphones and tablets increases, so does the development and use of dictionary apps. Gao (2013: 213) describes the present lexicographic situation as follows:

"In order to tap the potential of the vast global mobile market, dictionary publishers, large and small, have jumped on the appification bandwagon and launched their respective dictionary apps with the same zeal displayed a couple of years ago when they rolled out their online dictionaries."

Developing an app version of a printed or digital dictionary that meets the needs of both old and new user groups, however, is not that simple. During the development process, dictionary app developers must consider several fundamental lexicographic aspects. Surprisingly – and unfortunately for those who want to build on the experience amassed by other lexicographers – reports on the ideas and decisions behind each dictionary app are few in number (Holmer & Sköldberg, 2014; cf. e.g. Gao, 2013; Rundell, 2013; Simonsen, 2014a, b). Of course, all decisions made by lexicographers and app developers are based on ideas about target users and their specific needs. In that sense, the prerequisites and conditions for apps are diverse. Nevertheless, we believe there is a need for a wide-ranging discussion of the considerations that go into dictionary apps.

The principal aim of our paper is to further the increased exchange of views on, and experiences of, dictionary app development and usage. We will present the ideas, principles and the lexical database of the new app corresponding to the comprehensive printed dictionary, *Svensk ordbok utgiven av Svenska Akademien*, or SO (the Contemporary Dictionary of the Swedish Academy, 2009). SO contains about 65,000 headwords with thorough definitions. It includes, among other items, exhaustive pronunciation information and morphological information such as word parts, word formation and derivatives. The dictionary also provides about 25,000 etymologies, and the year of its first occurrence in Swedish is given for every numbered sense. There are also about 1,000 well-known literary citations and about 400 elaborated usage guidelines (see Malmgren, 2009; Malmgren & Sköldberg, 2013). The SO app is the result of close cooperation between the Swedish Academy, lexicographers and system developers at the Department of Swedish, University of Gothenburg (where the authors of this paper are employed), and Isolve AB, one of the leading dictionary app development agencies in Sweden.

In the next section we discuss dictionary app development and the results of studies on dictionary app use. Section 3 presents the lexical database and the IT environment at the Department of Swedish, University of Gothenburg. We discuss the SO app in section 4, with a focus on its content, design and search functions. Section 5 contains our final remarks.

2. Development and use of dictionary apps

A lexicographic team faces many issues when developing a dictionary app for smartphones and tablets. One important question concerns the app content in relation to the lexicographic data in the corresponding printed or online dictionary; that is, must the content be identical? Another key issue is how to display and make the most of the content while taking advantage of the inherent functionality of each platform. As Rundell (2013: 5) points out, a dictionary accessed on a computer or a mobile device has considerable advantages over its analogue predecessors. One obvious benefit is related to space. Gao (2013: 215) points out that "unlimited space offers the lexicographers a variety of choices, such as the addition of many entries, the multimedia content, the listing of related words, and the inclusion of more than one language in the dictionary, etc." However, according to Lew (in press) it is very important to make a distinction between *storage space* and *presentation space* in a lexicographic resource. Due to the size of a smartphone or tablet interface, the presentation space of dictionary apps is very limited and this must always be kept in mind when considering possibilities and preparing the data. But Lew also discusses a third category of space, which he calls *perceptual space*. This concerns the capacity of the dictionary user to perceive and process data. In other words, compared with storage and presentation space, perceptual space is not a property of the dictionary or the medium, but rather of the user. Lew states that presenting an overly rich microstructure can lead to information overload. He writes: "As a result, users find it difficult to extract the relevant information and may be less willing to proceed beyond the initial sense(s) of an entry" (see also Tarp, 2012 on problems related to information overload in dictionaries). Lew (in press) concludes that user research is needed to first establish what content should be immediately displayed on the screen, and what content should be deferred.

The issue of space has yet another aspect that is less often discussed by lexicographers. The possibility of accessing related content via hyperlinks in the text does in fact save storage space – memory – in any well-structured electronic dictionary because the need to duplicate information is more or less eradicated. In a printed dictionary, redundancy is necessary to avoid forcing the reader to shift focus from one entry to another; in an electronic dictionary, it is not only unnecessary but highly inadvisable. Duplication of information is the mother of inconsistency and should be avoided as far as possible, especially since the users of digital media often expect more frequent content updates, which serves to dramatically increase the problem of data integrity if the same information is stored in multiple locations.

There are also semi-technical decisions to be made when developing a dictionary app, such as whether the app is going to work online, offline, or perhaps be a hybrid of the two types. Among the dictionary apps developed in the Nordic countries, a clear majority seem to work offline, i.e., the entire dictionary content is downloaded to the phone/tablet upon installation. This applies for instance to the apps developed by Norstedts, the leading commercial dictionary publisher in Sweden. The dictionary apps developed by the Society for Danish Language and Literature, on the other hand, are online apps, which means that the mobile device must be connected to the internet to work. Merriam-Webster Dictionary apps can be classified as hybrids. No internet connection is required to view definitions and transliterations of pronunciation, however, users do need network access to hear audio pronunciations, study the illustrations and use the voice search feature. Generally, it can be regarded as a disadvantage if a mobile app requires network access since the connection might be slow, unstable, non-existent or expensive (Rundell, 2013: 5). However, the online format also has very clear advantages, not least in view of the possibility of linking to an online version of the dictionary, updating the content, and presenting an up-todate Word of the Day (see Holmer & Sköldberg, 2014).

Other issues arise when considering which mobile devices and operating systems to focus on targeting (iPhone, Android, etc.), and similarities and differences between operating systems, smartphone models and tablet versions (cf. Winestock & Jeong, 2014: 112).

Finally, a dictionary app producer must make decisions about pricing. Since the mid-2000s, dictionary sales have fallen sharply in Sweden. As Törnqvist (2010: 485–486) points out, many Swedish users now expect linguistic information to be available free of charge. Rundell (2013: 11) reports on the situation of English dictionaries, which appears more positive. However, he also states that the only digital products dictionary users seem to want to spend money on are those that can be installed on their own device, e.g. as an app.

Additionally, there is growing awareness of the importance of open access to data and software produced with the help of government grants, at least in the Nordic countries. For example, the Ministry of Education in Finland has demanded that any dictionary produced by the Institute for the Languages of Finland must be accessible to the public free of charge. This means that previously existing partnerships between academic institutions, commercial publishers and software producers must be reconsidered.

Depending on the commercial market for funding, dictionary producers have to think outside the box. Rundell (2013: 11) states that "What we need is a new entrepreneurship to create new products for new users, doing what we have always done: helping people to write, learn and understand language, working closely together with scientists and programmers to finally step into the digital future." Dictionaries are a fundamental component of the communicative and cultural infrastructure of a language community – the question is not whether dictionaries will exist in the future, but rather whether lexicographers will be a part of producing those dictionaries.

Very little information is available regarding the underlying ideas, visions and the actual development process behind the dictionary apps available today. Consequently, the usual procedure of benchmarking before embarking on a development project in order to grasp the state of the art is no easy task. There are few reviews of dictionary apps (see e.g. Holmer, 2011; Hoel, 2012; Svarverud, 2014 with a Nordic perspective). Holmer & Sköldberg (2014) examine four Danish apps developed by the Society for Danish Language and Literature. They conclude that these monolingual apps differ widely in terms of functionality and presentation of dictionary content. They also raise the issue of whether the app format is suitable for all kinds of dictionaries. The modern Danish Dictionary app (DDO) is perceived as a dynamic and very well-functioning lexicographic product that can serve as a model for other apps. The legacy dictionary apps (e.g. such as for the *Dictionary of the Danish Language*), are much simpler in terms of the online version of the same dictionary.

Holmer & Sköldberg (2014) also question to what extent dictionary app development is – or should be – based on the results of user studies. Surveys of dictionary app usage are still few in number. Marello (2014) compares the usage of three different versions of a bilingual dictionary (printed, online and app) among high school students. Simonsen (2014a, b) studies the usage of the dictionary app Medicin.dk, a knowledge-based medical resource used by most health care professionals in Denmark. Based on his data, Simonsen (2014b: 259–260) states that the typical mobile user is on the move and accesses information on the go, typically performing simple searches. This makes the mobile user impatient, imprecise and preoccupied with other things. The mobile user's situation primarily supports simple, precise, communicative lexicographic functions, and is not suited to support complex, cognitive lexicographic functions. Simonsen (2014a, b) also points out that the mobile user navigates both the physical world and the user interface of the mobile device at the same time. Finally, the size of the user interface and the typical user's situation mean that complex data and long text segments do not constitute optimal mobile data.

A user study that is highly relevant in relation to the development of the SO app is that carried out by the editors of the *Swedish Academy Glossary* (SAOL) (Holmer, Hult & Sköldberg, 2015). The SAOL is a monolingual Swedish glossary that contains about 125,000 headwords. It provides information primarily on spelling and inflection and explains the meaning of words to a minor extent. The app version of the SAOL was released in 2011 and is based on the 13th edition of the glossary, published in 2006. The app reflects the printed version and provides the full inflectional pattern for each lemma.

The app user study was performed in the early spring of 2015 in the form of a web survey. The questions concerned user behaviour and situations, opinions on the design and layout of the app, suggestions for a forthcoming version and background information about the respondents. The study resulted in 264 submitted questionnaires with a very low internal dropout rate.

The results of the study show that the app is mainly used for spelling, meaning, inflection and checking whether a particular word is included in the glossary. The respondents were fairly well-educated and often use the app in situations related to work. Overall, the respondents were very satisfied with the app and always or almost always find what they are looking for. When it comes to pricing and willingness to pay for a future version (the current version is free), older users were a bit more willing to pay, and most respondents said they would not want to pay more than 50 Swedish kronor (a bit more than 5 euros). In a forthcoming version, many respondents would like to have a wildcard search (a feature that is lacking in the current version), and cross-referencing via hyperlinks. By asking about their latest lookup, the editors discovered that the respondents tend to search for words that do not belong to the core vocabulary, and in their comments they explained that they were looking for object forms and variant forms of words, correct spelling, etc. Many of them also expressed a need for more detailed definitions, and some suggested an app version of the more comprehensive SO dictionary, i.e. the forthcoming app presented in this paper.

Before going into detail regarding the SO app, in the next section we give an overview of the lexical database and the related IT environment at the Centre for Lexicology and Lexicography, Department of Swedish, University of Gothenburg.

3. The Lexical Database and IT environment

3.1 Background

The origin of the database we are using to produce our app dates back to the mid-1960s when Sture Allén¹, a pioneer of computer-based lexicography at the University of Gothenburg, started gathering frequency-based data on contemporary language (see Malmgren & Sköldberg, 2013). These data evolved into a highly structured lexical database, designed mainly by Christian Sjögreen, a leading systems engineer at the subsequently formed Department of Computational Linguistics.



Figure 1: The Lexical Database, overview of processes, input and output

This database has since been continuously augmented and updated and used to produce a number of printed dictionaries in collaboration with different publishers (cf. overview in Figure 1). The latest printed dictionary produced was the SO, published in 2009. The database is currently owned by the Swedish Academy and maintained by the Department of Swedish at the University of Gothenburg. The Swedish Academy guarantees long-term funding of the work done by lexicographers, system administrators and developers at the university.

 $^{^1}$ See http://www.svenskaakademien.se/en/the_academy/members/938e01b1-b318-4c23-ba05-954127697d2a.

3.2 Infrastructure and traditional input/output

All data are stored in a dedicated relational database using the Ingres DBMS. With some exceptions, each information type has its own table (cf. Figure 2, where for example, information on pronunciation is stored in its own table with separate columns for primary and secondary pronunciation). Each main item has a unique number which acts as the key when joining tables, i.e. the classic relational database architecture.



Figure 2: Ingres tables and editing frame

The front end used for editing is designed in OpenRoad and users seldom need to be concerned with the underlying structure (see Figure 2). Editing is performed by scholars and PhD students specialized in lexicography. A number of in-house systems, such as corpora and morphological databases, have been developed in order to serve as resources for the lexicographic team.

Traditionally, output from the system has been through C programs producing LaTeX output converted either to PDF for human eyes or a format suitable for typesetting. Figure 3 shows an entry from the most recent printed dictionary (SO, 2009).

cam'pa äv. kam'pa verb ~de ~t + camp-ar, kamp-ar
sUBST.: campande, camping, kampande, kampning
övernatta i tält eller husvagn vanl. i samband med
turistande {JFR kampera}: de ~de på Västkusten i somras
KONSTR.: ~ (ngnstans)
HIST.: sedan 1925; av eng. camp med samma bet.; till lat.
cam'pus 'mark; fält; fältläger'; jfr ¹kamp

Figure 3: Dictionary entry for the word **campa** ('to camp') in SO (2009)

3.3 Redesigned for new media

In 2010 there were plans to create a Swedish language website presenting online versions of the dictionaries and other language resources owned by the Swedish Academy. Development addressing this goal resulted in a PHP/HTML application that was subsequently used as an in-house tool.

The main issues addressed during this phase were:

- Converting typesetting instructions into tag attributes + CSS style sheets
- Identifying and converting special characters into standard UTF-8
- Creating functions for identifying and linking referenced words

• Adapting the dictionary article layout to web browser functionality, e.g. using morphological information from another lexical resource² to produce flexible line breaks (soft hyphens)

- Amending inconsistencies and fixing referencing errors in the database
- Identifying parts of the text structure suitable for collapsing (see Figures 4 and 5)



Figures 4 and 5: HTML prototype showing collapsed text

² Svensk Morfologisk Databas (SMDB), an in-house tool for handling morphological data.

The decision was made to proceed with an app instead of a browser-based online version of the printed dictionary in 2013. App production was contracted to Isolve AB. The content of the app was produced using a modified version of the existing PHP/HTML application. The XML file exchange format was used for enhanced verifiability and the structure was formalized in an XSD schema file (cf. Figures 6 and 7). Audio files were also added.

	_
<pre><artikel><lemma id="lnr129692"><homonr>cam`pa<ljudfil filnamn="129692_1.mp3"></ljudfil><alt><artikel></artikel></alt></homonr></lemma></artikel></pre>	
<td></td>	
<pre>id="xnr129503"><kernel id="knr450910"></kernel>cpunkt><def>wr129503"><kernel id="knr450910"></kernel>cpunkt><def> <deft>waligen i sam-band med turistande</deft> <fack>sport.</fack></def></def></pre>	
<hv><hvgrupp><hvtyp>JFR</hvtyp><hvtyp2>cohyponym</hvtyp2> <hvtag>xnr202494</hvtag><hvord><hvhomo></hvhomo>kampera</hvord></hvgrupp></hv>	
<pre><detaljer id="detaljer129693"><exempelblock><sxblocklx><syntex>de campade på Västkusten i somras</syntex></sxblocklx></exempelblock> </detaljer></pre>	1
<pre>cvarehs<vtcs.mba (ightstars)="" <="" vt="">cvarehs</vtcs.mba></pre> cvarehs	
129692 <hvord><hvhomo></hvhomo>campande</hvord> , <ordbildn>hvtag>Inr129696<hvord><hvhomo></hvhomo>camping</hvord>, </ordbildn> <ordbildn>hvid>vbid3-129692<hvord><hvhomo></hvhomo>kampande</hvord>, </ordbildn> sordbildn>hvid>vbid3-	
129692 <hvord><hvhomo></hvhomo>kampning</hvord> 	





Figure 7: XML schema description

4. The SO dictionary app

It is a challenge to develop a dictionary app that not only accurately reflects the comprehensive printed dictionary from 2009 but that can also be regarded as an independent lexicographical resource. Release of the app is planned for late summer of 2015. As we have indicated, the app will be the result of close collaboration and a

working method characterized by flexibility, especially between the lexicographers and system developers in Gothenburg and the app developers at Isolve AB in Stockholm. During the process, errors in the database and extraction programs have been identified and fixed and different app versions have been examined and tested by an extensive test group consisting mainly of lexicographers at the University of Gothenburg and employees at the Swedish Academy. The app has not yet been subjected to a user study. However, the editors and system developers have been able to draw some conclusions about app user behaviour thanks to the recently performed user study on the related project, the *Swedish Academy Glossary*, or SAOL (see section 2).

The primary target user groups of the printed version are native Swedish speakers and advanced learners. The dictionary is polyfunctional; it supports both receptive and productive user situations, while also fulfilling a documentary function. It has not yet undergone a user survey, but according to letters to the editorial board, it is mainly used by translators, proofreaders, language teachers, etc., i.e. people who work with the Swedish language on a professional level. However, as a result of the app format, the dictionary has the potential to reach a wider user group, for example younger and less experienced dictionary users. Taking this into account, it is very important to avoid information overload; the lexical data must be presented in a way the users can handle, otherwise it might hinder the retrieval of information needed (cf. Tarp, 2012: 255 and Lew, in section 2).

4.1. Platforms

The SO app is designed for iOS and Android. This decision was based on the dominant position of the iPhone and Android operating systems in the Swedish mobile market. Furthermore, statistics on the number of downloads of the dictionary app of the SAOL glossary show that those platforms are the most common among Swedish users. The results of the survey of active SAOL app users also support this idea.

The SO app is a hybrid of an online and offline resource. Although the app allows users to look up words and to view definitions and examples offline, a network connection is required to access audio pronunciations. The main content of the app is static, but some of the information can be updated as soon as an internet connection is available, e.g. the sections concerning the Swedish Academy, related apps, etc.

4.2 User interface – layout

The user interface described here is the current version of the iPhone GUI. The iPad GUI is similar to the iPhone GUI, but we take advantage of the larger screen size. For example, the list of entries and the current entry are shown on the same screen on iPads. When this paper was written, development of the Android version of the app had only just begun.

The main screen of the app contains the status bar, a navigation control, a search bar, a few standard icons and a text field (the content of which depends on the current activity). The user can choose between searching for a word $(S\ddot{o}k)$, bookmarks $(Bokm\ddot{a}rken)$, search history (*Historik*), usage guidelines for particular words (Stilrutor), Word of the Day (*Dagens ord*), news (*Aktuellt*), general information about the app, the Swedish Academy, help, etc. and giving feedback (*Återkoppling*).

The default mode is Search, and the user is presented with a list of dictionary entries beginning with A when the app opens.

Compared with other dictionary apps, such as those developed by the Society for Danish Language and Literature, a considerable amount of information is provided concerning the dictionary content. The user instructions are also relatively comprehensive. According to Svensén (2009: 459), "it is a truth universally acknowledged in lexicographic circles that user's guides are very seldom consulted", but the user survey on the SAOL app shows that a relatively high number of the respondents consulted this information in the glossary app. With the aim of developing the SO app to be as independent as possible from its analogue predecessor, we find it important for users to be able to get all of the information on lexicographic content in the app without being forced to consult the book.



Figure 8: Drawer menu in the SO dictionary app

The *Stilrutor* ('style boxes') screen shows a list of all dictionary entries that include usage guidelines. For example, in the list you can find the entry *genitiv* ('genitive') followed by instructions on correct usage of genitive apostrophes in Swedish. In comparison with the printed dictionary, these guidelines have enhanced visibility and are more easily accessed in the app, since they can be found not only when you happen to look up a particular word, but also through the action item *Stilrutor*.

In the *Historik* screen ('search history'), shortcuts to recent lookups are listed chronologically. The *Bokmärken* ('bookmarks') screen contains a similar list of shortcuts to bookmarked dictionary entries.

The intention behind introducing headwords as *Dagens ord* ('Word of the Day') is to provide a sample of the comprehensive content of SO. It is hoped that the selected entries will serve as illustrative examples of entries and pique the user's interest in delving more deeply into the dictionary content. They may also stimulate vocabulary building and support language acquisition, especially among learners of Swedish. In some dictionary apps, such as Dictionary.com, the *Word of the Day* seems to be selected at random. In the online working Danish DDO app, the headword is continuously refreshed, resulting in a dynamic, fresh "look" (see Holmer & Sköldberg, 2014).

In the SO app, the Word of the Day display does not require internet access. The selection of words is made in advance by the lexicographers. A list of entries is prepared for a period of one year and the entries are set for specific dates. Some of the selected entries are closely connected to a certain season (e.g. krokus, 'crocus') or a Swedish feast day (e.g. påskägg, 'Easter egg'). Furthermore, some words on the list derive from Old Swedish, e.g. the noun dag ('day'), which dates from the 9th century, and others are relatively new loanwords (e.g. sudoku 'sudoku'). Many are also conspicuously metaphorical (e.g. flaskhals, 'bottleneck', bokslukare, 'book swallower', i.e. a voracious reader). With this feature, we thus hope the Word of the Day in the SO app will increase interest in the Swedish language and its vocabulary in general.

The Swedish verb *campa* ('to camp') is presented in Figure 9.

●○○○○ Telenor SE 🗢	16:34	1 72 % 💷	
=	Sök		
Q campa	8	Avbryt	
campa			
campare			
camping			
campingplats			
campus			
canacta			
QWER1	YUI	ЭРÅ	
ASDFO	ЭНЈК	LÖÄ	
• z x c	VBN		
123 🌐 m	nellanslag	Sök	

• ○ ○ ○ Telenor SE
🔍 campa 🛛 😒
 cam`pa [♠] äv. kam`pa verb ~de ~t oRDLED: camp·ar, kamp·ar övernatta i tält eller hus- vagn vanligen i samband med turistande JFR kampera EXEMPEL: de campade på Västkusten i somras KONSTR.: campa (ngnstans) HIST: sedan 1925; av eng. camp med samma betydelse; till lat. cam´pus'mark; fält; fältläger'; jfr ¹kamp
subst.: campande, camping, kampande, kampning



As shown in Figure 9, the user interface of the app is mainly blue, white and black. The headwords are presented in red to make them stand out. We intend for the interface design to be perceived as stylistically clean and aesthetically pleasing. The aim of the design is also to connect the app to the deep blue cover of the printed version of the SO and thus make the app seem familiar to people who have previously used the printed dictionary. At the same time, the connection with the related app, the SAOL (also financed by the Swedish Academy), must be evident. In light of these facts, the icon, which is blue and includes the classic Academy emblem – a laurel wreath surrounding the Academy motto "Snille och smak" ('Talent and Taste') – was chosen.

The default text size in the SO app is comparable to the text size in other dictionary apps. Hopefully users will successfully hit the touch zones on the screen and the keyboard even when on the move, which is not unusual in app usage (see Simonsen in section 2). In order to meet the needs of different users and user situations, the display of the article content is adapted to a smaller or bigger font size in response to standard pinch-in/pinch-out zooming gestures. This is made possible in the SO app thanks to built-in soft hyphens, which allow for dynamic word wrapping (see section 3).

An additional aspect of the app design is that users can switch between portrait and landscape modes simply by turning the mobile or tablet (cf. for example, dictionary apps by Longman and Merriam-Webster in this respect). Even though users have personal preferences for portrait or landscape, their choices are also affected by their situations. For example, people may tend to watch video in landscape mode and read in portrait. It is important for users to be able to make their own choice and individualize their usage. In relation to this function, the use of soft hyphens is also important; without them, the right margin of entries could appear ragged.

4.3. Content and search functions

The entire lexicographic content of the printed version of the SO is included in the app. In this regard, the SO app is quite different from the Danish DDO app, which only provides a sample of the content found in the web version (see Holmer & Sköldberg, 2014). If users wish to see everything in that lexicographic resource, they are obliged to consult the dictionary site ordnet.dk (easily accessed via links in the app). It could be argued that the links in the DDO encourage users to go to the online version of the dictionary. As a result, the app could be regarded as a spin-off of the web version. But, as previously mentioned, the DDO app is a highly effective lexicographic product, so this is not the case. A similar model in the SO app with external links to an online version is not possible because the Swedish Academy decided to prioritize the dictionary app rather than an online version (see section 3).

An addition to the lexicographical content of the SO is the inclusion of approximately

65,000 human-read audio files. These files constitute an important aid for users, especially learners of Swedish. Access to audio pronunciation will probably also increase interest in the dictionary among native Swedish speakers. However, the integration of audio pronunciation also raises new issues. When users can listen to pronunciations of the headwords, will phonetic transcription – information that many users have difficulties interpreting without consulting the pronunciation key – still be necessary? (cf. Svensén, 2009: 383). In such an online resource there is plenty of presentation space; but presentation space is very restricted on a mobile screen (see Lew in section 2). We have decided to keep the phonetic transcriptions in the dictionary app for two main reasons. As Lew (in press) points out, learners of Swedish may not be able to hear phonemic distinctions since their perception is filtered through the phonological system of their native language. Moreover, as mentioned, audio pronunciation is only accessible when the user is connected to an internet network.

A common, simple search is performed by starting to type the sought-for word. The list of matching entries adjusts as the user types. The headword is shown at the top of the list followed by the rest of the dictionary headwords (compared to, for example, the DDO app, which only shows the next 29 headwords). Like in most dictionary apps, it is possible to scroll up and down in the lemma list. This function is essential for people who want to gain an understanding of nearby headwords, something that is of course simple in a book. By clicking a lemma in the list, the whole entry is shown. The search algorithm also supports searches for inflected forms (algorithm developed by Isolve AB).

Users can additionally perform phrase searches. The algorithm is the same as for a simple search. The search string "kalla fötter" ('cold feet') generates the idiom variants fa^{*} kalla fötter ('get cold feet') and ge ngn kalla fötter ('give someone cold feet') (see Figure 10). But the search string also generates other results from other entries in the dictionary containing the word forms cold and foot, for example, a syntactical example in the entry doppa ('dip'): "The water was so cold she only dipped her feet". The word forms in the search string are distinguished in the hits with bold typeface. In each example, information on the entry (in blue) and the information category (in grey) is given. In this particular case, users are informed that the idiom fa^{*} kalla fötter is placed under the noun fot ('foot'). They can also easily check the entry in question, as it constitutes a cross-reference. The lexicographers decided which types of phrases were to be indexed and used for this search. In short, idioms and other fixed phrases that include two or more word forms given in the search string are presented at the top of the list because it is reasonable to assume that this is the multi-lexical unit that users want in most cases.



Figure 10: Result of a phrase search in the SO dictionary app

There is also a spell-check function developed by Isolve AB.

The SO app also supports wildcard search. A search string like "*boll*" ('*ball*') generates hits such as *bollhav* ('ball pit'), *fotboll* ('football') and *snöbollseffekt* ('snowball effect'). These kinds of searches may appeal to scholars. Considering that there is no online version of the SO, the app may be used to perform different kinds of lexicological studies. This function may also very well appeal to users interested in solving e.g. crosswords.

When it comes to article microstructure, the users of the printed SO will probably find the layout familiar. The italics and different type sizes are still there. However, the printed version of the SO, like many other dictionaries, is characterized by compression (cf. Lew, in press). With the aim of making the entries and information more accessible to users, more headings are included and many of the abbreviations are dissolved and shown in full text, for example, for the part of speech, the tilde used to mark the lemma in the entry text is replaced with the lemma, etc. Even though the display area on a mobile device is very limited, we find this an important consideration by the users, especially learners of Swedish.

4.4. Collapsing and expanding

To make extensive dictionary articles clearer and the dictionary content easier to grasp, longer entries in the SO app are shown in a collapsed form. See Figure 11 for examples of a collapsed and an expanded version of the noun *harmoni* ('harmony').



Figure 11: Two versions of the entry **harmoni** ('harmony'), collapsed version (left) and extended version (right).

A relevant question is what is considered to be a "long" or "short" entry (cf. Trap-Jensen, 2010). We have chosen to collapse dictionary entries that display over more than one screen size (iPhone 5). In the development process we have been experimenting with the optimal amount of data presented by default using the HTML prototype (see Figures 4 and 5). Tarp (2012) states that the problem concerning individual entries on the screen is not only how much *can* be presented at a given time to a dictionary user, but also how much should be presented. In the present app version, details that belong to separate core meanings are hidden if all data cannot fit onto one screen so that the user gets a clear overview of the semantic structure. We present the following information categories in the collapsed view: headword, pronunciation, part of speech, inflected forms, definitions (of core meanings) and related words (like synonyms and antonyms). By touching the expansion symbol (the plus sign), users can access subordinated meanings, idioms, information on valency, etymology, etc. Lew (in press) concludes that user research is needed to establish what content should be displayed immediately on the screen, and what content should be deferred. We hope to perform such a study when the app has been on the market for a while.

4.5 Cross references and hyperlinks

The printed dictionary contains a considerable number of references to specific meanings of related words. For comparison, Figure 11 shows links in blue. These references have been implemented as hyperlinks in the app by means of supplementing XML tags with ID/IDREF attributes. During this process, a number of faults in the database were brought to light, such as references to words that had been excluded in print. Finding these kinds of errors is common to any IT

development project and must be taken into account when embarking on an appification project.

5. Final remarks

In this paper we present the ideas behind a new Swedish dictionary app, which we hope will reflect the comprehensive *Contemporary Dictionary of the Swedish Academy* (the SO), 2009. We present the lexical database that has been evolving since the mid-1960s and which has resulted in numerous scientific reports, printed dictionaries, internal web interfaces and finally a dictionary app. We also highlight strategic considerations for optimising the layout and presentation of the database content so that it fits the app display while retaining as much as possible the look and feel of the physical book.

The SO app will cost 49 Swedish kronor (about 5 euros), which is competitive compared to the printed dictionary, which costs about 500 Swedish kronor (a bit more than 50 euros). The price of dictionary apps on the Swedish market, such as Norstedts, range from 49 Swedish kronor for the smaller ones to 390 Swedish kronor (40 euros) for the most comprehensive bilingual Swedish/English dictionary. Thus, the SO app can be considered heavily subsidized. It is well-known to lexicographers that dictionary projects are expensive, take a long time and are never really finished. But as the SAOL app user study has shown, many users are not really willing to pay for dictionary apps, and even if they are, they are not prepared to pay very much. Even though the Swedish Academy could in theory give away the app for free, taking the decision to charge a small amount reflects a desired position concerning high quality lexicographical products.

This article aimed to participate in a broader discussion of experiences involved with producing dictionary apps, app development and app user behaviour. In this paper, we focussed on the mobile phone app since the tablet app is somewhat different. The app presented here is planned to be released in late summer of 2015. This app has been tested by an extended test group, but has not yet been the object of a user study *per se.* So far, the SO printed dictionary has not been researched from a user's perspective either. However, the editors and system developers were able to draw some conclusions regarding app user behaviour as a result of a user study on a related project, the *Swedish Academy Glossary* (SAOL) that was just carried out in March 2015.

It is possible to approach the use of online dictionary apps with log files and statistics. App developers who want to gain insight into user behaviour with offline dictionary apps may be supported by mobile app measurement and advertising platforms like Flurry Analytics from Yahoo! (http://www.flurry.com/). By implementing Flurry in the SO app, the lexicographic team and app developers can gain deeper understanding of app user behaviour through the analysis of usage data, such as lookups, session duration, operative systems, device models, etc. Flurry will

be implemented in the SO app and knowledge about how the app is actually used will be invaluable when preparing updates and improving future versions.

6. References

Dictionary.com. Dictionary.com, LCC. App version 5.2.1. (Accessed 25 May 2015)

- DDO: *Den Danske Ordbog.* Det Danske Sprog- og Litteraturselskab. App version 2.0.11 (iOS). (Accessed 25 May 2015)
- Gao, Y. (2013). The Appification of Dictionaries: From a Chinese Perspective. In Kosem et al. (eds.), *Electronic lexicography in the 21st century: thinking outside* the paper. Proceedings of the eLex 2013 conference. Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, pp. 213–224.
- Hoel, J. (2012). Appsolutt fingerferdig! En anmeldelse av ordbokappene RO og SAOL. *LexicoNordica* 19, pp. 255–271.
- Holmer, L. (2011). Norstedts ordboksappar. LexicoNordica 18, pp. 307–322.
- Holmer, L, Hult, A.-K. & Sköldberg, E. (2015). Spell-checking on the fly? On the use of a Swedish dictionary app. Proceedings of eLex conference 11-13 Aug. 2015.
- Holmer, L. & Sköldberg, E. (2014). Appifiering till allas lycka? Om danska ordboksappar med särskilt fokus på DDO. *LexicoNordica* 21, pp. 235–252.
- Lew, R. (in press). Space restrictions in paper and electronic dictionaries and their implications for the design of production dictionaries. In P. Bański & B. Wójtowicz (eds.) Issues in Modern Lexicography. München: Lincom Europa.
- Malmgren, S.-G. (2009). On production-oriented information in Swedish monolingual defining dictionaries. In: S. Nielsen & S. Tarp (eds.) Lexicography in the 21st century. In honour of Henning Bergenholtz. Amsterdam/Philadelphia: John Benjamins, pp. 93–102.
- Malmgren, S.-G. & Sköldberg, E. (2013). The lexicography of Swedish and other Scandinavian languages. International Journal of Lexicography 26(2), pp. 117– 134.
- Marello, C. (2014). Using Mobile Bilingual Dictionaries in an EFL Class. In A. Abel,
 C. Vettori & N. Ralli (eds.) Proceedings of the XVI EURALEX International Congress: The User in Focus. 15–19 July 2014. Bolzano/Bozen, pp. 63–83.
- Merriam-Webster: Merriam-Webster Dictionary. Merriam-Webster Inc. App version 3.2 (iOS). (Accessed 25 May 2015)
- Rundell, M. (2013). Redefining the dictionary: From print to digital. In *Kernerman Dictionary News* 21. Available at: http://kdictionaries.com/kdn/kdn21.pdf.
- Simonsen, H. Køhler (2014a). Brugerne er allerede mobile! In R. Vatvedt Fjeld & M. Hovdenak (eds.) Nordiske studier i leksikografi 12. Oslo: Novus, pp. 416–429.
- Simonsen, H. Køhler (2014b). Mobile Lexicography: A Survey of the Mobile User Situation. In A. Abel, C. Vettori & N. Ralli (eds.) Proceedings of the XVI EURALEX International Congress: The User in Focus. 15-19 July 2014. Bolzano/Bozen, pp. 249–261.
- SAOL13: The Swedish Academy Glossary. (2006). 13th edition. Svenska Akademien &

Isolve AB. App version 1.1.8 (iOS). (Accessed 25 May 2015)

SO: Svensk ordbok utgiven av Svenska Akademien. (2009). Stockholm: Norstedts.

- Svarverud, R. (2014). Nye kvalitetsverktøy for brukere av kinesisk i Skandinavia. LexicoNordica 21, pp. 341–356.
- Svensén, B. (2009). A Handbook of Lexicography. The Theory and Practice of Dictionary-Making. Cambridge: Cambridge University Press.
- Trap-Jensen, L. (2010). One, Two, Many: Customization and User Profiles in Internet Dictionaries. In A. Dykstra & T. Schoonheim (eds.) Proceedings of the XIV Euralex International Congress. Leeuwarden: Fryske Akademy, pp. 1133–1143.
- Tarp, S. (2012). Online dictionaries: today and tomorrow. Lexicographica 28, pp. 253– 267.
- Törnqvist, L. (2010). Ordböcker på Internet och Internet som ordbok. In H. Lönnroth & K. Nikula (eds.) Nordiska studier i lexikografi 10. Tammerfors, pp. 484–493.
- Winestock, C. & Y.-k. Jeong (2014). An analysis of the smartphone dictionary app market. *Lexicography Journal of ASIALEX* (2014(1), pp. 109–119.

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

http://creativecommons.org/licenses/by-sa/4.0/

