

Christian Mattsson

The Swedish Lexin Project – dictionaries in 22 minority languages

Lexin is a series of dictionaries in 22 minority languages, especially targeted at immigrants. This article is a presentation of the Lexin project with focus on the dictionaries published on the Internet.¹ A more theoretical account of Lexin is given by Gellerstam (1999), and a critical review by Loenheim (2012).

1. Lexin and the Swedish Language Act

Let's start by looking at a passage of the Swedish Language Act (2009:600) which was adopted in 2009. The purpose of this act is to specify the position and use of Swedish and other languages in Sweden. The act is also intended to protect the Swedish language and language diversity in Sweden, and the individual's access to language:

All residents of Sweden are to be given the opportunity to learn, develop and use Swedish. In addition, persons belonging to a national minority are to be given the opportunity to learn, develop and use the minority language [...] Persons whose mother tongue is not one of the languages specified² [...] are to be given the opportunity to develop and use their mother tongue.

This is exactly what Lexin sets out to do: to help people to develop not only their Swedish language skill, but also to develop their skills in their own native language. So when the first Lexin dictionaries were published in the early 1980s you could say that the lexicographers were very much ahead of their time, of rather ahead of legislation. Making dictionaries for minorities is quite in line with the Swedish Language Act.

2. Dictionaries in 22 languages

The Lexin project has grown continually since the beginnings about 30 years ago, and today we have dictionaries in 22 minority languages. These are *Albanian, Amahric, Arabic, Assyrian, Azerbaijani, Bosnian, Croatian, Finnish, Greek, Macedonian, North Kurdish (Kurmanji), Pashto, Persian, Romani, Rumanian, Russian, Serbian, Somali, South Kurdish (Sorani), Spanish, Tigrinja* and *Vietnamese*.

Sweden used to be more or less monolingual but this is no longer the case: today we have about 200 minority languages in Sweden. In fact, it is not easy to keep track of all languages spoken in Sweden. Unfortunately, we don't have any official statistics to tell us how many speakers there are of a certain language. When you come to Sweden as an immigrant you are not asked what language you speak, just your citizenship. In consequence, it can sometimes be difficult to know what languages need a dictionary the most.

¹ <http://www.lexinlexikon.se>.

² I.e. not Finnish, Meänkieli, Romany, Sami and Yiddish.

Of these estimated 200 minority languages Lexin includes only 22. This might seem a fairly limited selection, but we believe that these 22 languages meet the needs of the vast majority of immigrants in Sweden. In addition, some languages, e.g. Russian and Arabic, can often be used as a second language by people who don't have access to a Lexin dictionary in their own native language.

The majority of the Lexin dictionaries have some 28,000 entries including idioms and examples of language in use. Some have less, but we hope to extend these in the future. Most of the dictionaries are now available on the Internet. However, some of the older dictionaries are only available in print. Our website has proved to be a great success, and in 2012 we registered more than 170 million searches, which equals more than 5 searches every second. This is a 25 per cent increase since 2011.

3. From a monolingual into a bilingual dictionary

Lexin is basically a Swedish monolingual dictionary which has been translated into 22 different minority languages. The printed dictionaries all go *from* Swedish into a foreign language, so there are only Swedish-Arabic, Swedish-Albanian dictionaries etc., not in the other direction. However, on the Internet you can look up a word in the target language as well. But what is shown is the Swedish entry where the translation in the target language occurs. This is true *only* if this word is included as a translation of a Swedish word. In consequence, there are no culture specific words of the target language included, simply because they don't appear as translations of a Swedish word. But this rarely seems to be a problem for the users, since we have had very few complaints.

The reason for not making genuinely bidirectional dictionaries is mainly financial. One could also argue that it is more important for the users to be able to decode Swedish, and to understand a Swedish word, rather than to find out how a word in your own language is translated into Swedish. So in this respect the dictionaries online are more useful than the ones in print. In addition, access to our Internet site is free, of course, whereas you have to pay for the printed dictionaries.

4. Selection of entries

In addition to the 28,000 or so most frequent words and phrases, which can be said to make up the basic vocabulary of the Swedish language, Lexin also lists a number of words that describe the Swedish system, i.e. the workings of the Swedish society, concepts which have to do with the health care system, pensions, sick leave, laws and regulations, organizations, political parties etc. As a newcomer to Sweden you are not only confronted with a strange and difficult new language, you are also immersed in a whole new society and culture, which can be more or less incomprehensible at first. So these words are essential in a dictionary for new citizens, but seldom found in other dictionaries.

5. The dictionary entries

The dictionaries contain a wealth of information. Let's look at the entry *hand* (which is the same word in both Swedish and English) in the Albanian dictionary:

☐ **hand** [han:d] LYSSNA subst.
 ⟨hand, handen, händer⟩
 yttersta delen av en människas arm

dorë
 BILD SVENSKA / BILD ALBANSKA

Exempel:

- tvätta händerna – pastroj duart, laj duart
- sy för hand – qep me dorë
- gå hand i hand – eci dorë për dorë

Uttryck:

- ta hand om ("vårda o dyl") – kujdesem për diçka (dikë)
- efter hand ("fortlöpande") – shkallë-shkallë, gradualisht
- på egen hand ("själv") – vetë, në mënyrë të pavarur
- i första hand ("viktigast, främst") – pikësëpari, së pari, para së gjithash
- ha till hands ("ha klart till användning") – kam gati, kam në dorë, zotëroj

Sammansättningar:

- hand|bagage – bagazh dore

Figure 1: The entry “hand” in the Swedish-Albanian dictionary

The information categories include the following: Pronunciation is given both in writing and as recorded sound files. Part of speech is indicated throughout, as well as inflexion.

A feature unique to Lexin is that the present tense, not the infinitive form, is used as the entry form for verbs. The reason is pedagogical: in Swedish it is easier to infer the other forms of a verb if you look at the verb in present tense, rather than the infinitive form.

When it comes to the semantic part we are somewhat unorthodox, in that we present not only the translation in the majority language, but also the definition of the Swedish word. So here it is obvious that Lexin is both monolingual and bilingual at the same time. This doubled information can be very useful, since many Lexin languages are culturally more or less different from Swedish, indeed many are non-European. This means that some words or concepts have no exact match in the target language. So the user can get help from the translation, which can be more or less approximate at times, as well as from the more exact definition.

Another feature that is particularly useful for learners is that all words are clickable. This means that the user can click on any word in a translation or in a definition that she or he doesn't understand and be directed to the entry where this word is explained or translated.

The Lexin dictionaries also contain plenty of examples of usage, the contexts in which the word is typically used, both semantically and syntactically. Here the user is guided

e.g. on the use of prepositions. Lexin also lists many idioms and expressions, very much appreciated by the users, since this is often one of the most demanding parts of the language for a foreigner speaker.

Many entries also include a section of compounds, which is a very common feature in Swedish. In fact, compounds can be difficult to find in a dictionary if you don't even know if the word *is* a compound or written in two words. So when you look up e.g. the word *hand*, you can see all other entries where the word *hand* occurs, both as part of a compound and as a single word. Many entries also have syntactical information, which is intended mainly for more advanced users. So we don't expect all users to make use of all the details in the dictionary. Indeed, the users can be anything from almost illiterate to people with a Ph.D.

6. Illustrations and videos

However, Lexin is not all about words. Some of the most basic concepts are perhaps better explained by a picture. And so we have included some 1,800 illustrations, mainly of nouns, which all have been grouped together thematically. For instance the Swedish word *snö* (snow) is included in the theme called weather, together with *rain*, *hail*, *storm* etc.

When it comes to verbs a picture may seem somewhat static. Here we have attached over 100 short videos which illustrate all sorts of basic human activity, e.g. *yawn*, *laugh*, *cry*, *run*, *cut* and *sleep*. Both the pictures and the videos are particularly useful as a starting point for those with very little training in their own language. In fact, there are many immigrants who can neither read nor write in their own language when they arrive in Sweden. And of course children too appreciate this.

7. Sharing experiences – encouraging further development

We believe that Lexin dictionaries have made integration easier for all those who come to Sweden from all over the world, and now make up more than 10 per cent of our population. And Lexin has proved to be a useful model for dictionary making not only in Sweden. The other Scandinavian countries too have been inspired by Lexin, above all Norway, where Lexin started in the mid 90s. To date dictionaries in 15 minority languages have been produced, many of which have been adapted from the Swedish data base. This was made possible by an agreement giving Norway free access to use our dictionaries to make their own. Of course the adaption from Swedish to Norwegian was greatly facilitated by the fact that these two languages are very closely related. Actually it was possible to recycle or reuse about 80 per cent of the original Swedish bilingual dictionaries. South Africa too has shown interest in Lexin lately. This country has eleven official languages and there are tentative plans for a monolingual English dictionary which will then be translated into all eleven languages, using Lexin as a model.

We are very happy to share our experiences, but also our data bases. With the growing demand for free language tools and linguistic infrastructure we have decided to make all our dictionaries available under the Creative commons license (BY). This means that anyone can reuse our data to create other dictionaries or applications, teaching aids etc.

We also allow commercial use as long as the origin of the dictionary data is indicated. In this way we can contribute to a better linguistic infrastructure, by making quality dictionaries available for anyone to develop. We believe that this is a good way to promote the Swedish language.

8. References

Creative Commons licenses: <http://creativecommons.org>.

Gellerstam, M. (1999): Lexin – lexikon för invandrare. In: *LexicoNordica* 6, 3-18.

Lexin online: <http://www.lexinlexikon.se>.

Lexin online (Norwegian website): <http://decentius.hit.uib.no/lexin.html>.

Loenheim, L. (2012): Lexins svenska lexikon. In: *LexicoNordica* 19, 295-316.

Swedish Language Act: <http://www.government.se/sb/d/8317/a/138133>.

