The Authentic Voices of Dictionary Users – Viewing Comments on an Online Learner's Dictionary Before and After Revision

Ann-Kristin Hult
Department of Swedish, University of Gothenburg
ann-kristin.hult@svenska.gu.se

Abstract

This paper deals with comment field data from two web questionnaire surveys, performed in 2007 and 2011, about the use and users of the *Swedish Lexin Dictionary* (SLD), an online monolingual learner's dictionary. The SLD underwent comprehensive revision in between the two studies. In order to evaluate the update I compare the respondents' comments on SLD before and after the revision. The two sets of comment field data are categorised by the types of comments expressed and by the information categories mentioned. As it turns out, respondents do seem happier with the new version of SLD. Generally, there are more positive comments after the revision than before and the information categories mentioned are more often set in positive contexts in the latter data. In addition, the majority of respondents in 2011 belong to the dictionary's target group, which was not the case in 2007. Although respondents seem more satisfied with the new version of SLD, the two sets of comment field data contain largely the same kind of criticism, as for example in comments concerning (usually the lack of) lemmas and examples.

Keywords: dictionary use; learner´s dictionary; user comments

1   Introduction

Comment fields are often provided in questionnaires to give respondents the opportunity to expand upon what they have already reported. This article deals with comment field data from two web questionnaire surveys, performed in 2007 and 2011, about the use and users of the online monolingual *Swedish Lexin Dictionary*, SLD (http://lexin2.nada.kth.se/lexin/) (Hult 2008a; 2012). In the interval between the two studies the SLD underwent a comprehensive revision. One way of evaluating the success of the updating project is to compare respondents' comments on SLD before and after the revision. Considering the improvements made, are the respondents in the 2011 study happier dictionary users than those in the 2007 study? In an attempt to answer this question I will first briefly describe the updating project. I will then categorise the two sets of comment field data according to the types of comments expressed. Thereafter, I will examine what information categories are mentioned in the

comments, and what is said about them. Finally, I will give voice to the users by quoting from the comment field data, focusing on the "Lemma", "Meaning" and "Example" categories.

In the age of electronic dictionaries, users can and do play a much more active role in the process of dictionary making. Lexicographers may disagree on exactly how active users should be in this process (cf. Lew, in press). Be that as it may, since users are the consumers of our products, the least we can do is listen to what they have to say about them.

2   The Swedish Lexin Dictionary and the Updating Project

Lexin, short for "lexicon for immigrants", was originally a series of paper dictionaries at the advanced beginner's level for users with Swedish as a second language. Lexin includes one monolingual Swedish part, the SLD, which is the focus of this article, and about twenty bilingual dictionaries between Swedish and non-Scandinavian immigrant languages in Sweden such as Arabic, Russian and Somali. The Swedish material served as a basis for the bilingual dictionaries. The dictionaries are primarily aimed at recent immigrants to Sweden who are just beginning to learn Swedish. They are intended to be easy to use even by people with limited reading ability and little or no experience using dictionaries. Consequently, in comparison with many other dictionaries, a user-friendly layout is extremely important (Gellerstam 1999: 5). The SLD (and the entire Lexin series) is primarily intended for reception and secondarily for production. The lemma selection has been carefully adjusted to the needs of the dictionary's target group. All lemmas have been augmented with comprehensive information typical of a general dictionary. In addition, SLD includes specific words covering special issues and institutions in Swedish society, along with factual information. About 1,800 elementary words are illustrated with pictures in a special section that covers 31 themes, such as *The Human Body (external and internal)* and *Cooking and Meals.*

The free online version of SLD and the Lexin series have been available since 1994. In addition to pictures, the online version features animations that demonstrate the meanings of 700 verbs. These are classified in fifteen main sections, such as *The Kitchen and Cooking, Emergency and Medical Service* and *Travel and Transport (City and Traffic).* The animations make verbs otherwise difficult to make comprehensible through pictures – let alone definitions – much easier to understand. There are links from the dictionary entries to the pictures and animations. Use of the online version has steadily increased from the start. The site presently has about 20 million searches per month and the SLD has slightly more than one million searches per month (http://lexin2.nada.kth.se/statistik/html). The recent revision of the SLD encompassed both dictionary content and the website interface. Eventually, the bilingual dictionaries will also be revised.

The revision focused on strengthening the nature of the SLD as a learner's dictionary, especially its function as a reception dictionary (Malmgren 2012: 456). It is the dictionary's first comprehensive revision since it was released some thirty years ago. Naturally, one important task was to update the

lemma selection, which increased by about 3,000 lemmas. A great many synonyms and antonyms were also added. This was done mainly to improve users' chances of understanding the entry words: "one single synonym – even if not a perfect one – can make the difference between understanding and not understanding" (Hult et al. 2010: 807). The synonyms and antonyms also serve the purpose of increasing the users' vocabularies. Moreover, the update included a thorough adjustment subdividing the senses of each lemma and placing the examples in the immediate vicinity of the sense being illustrated. A great deal of effort was also dedicated to more explicitly describing particle verbs, which are often stumbling blocks for learners of Swedish. The structure of the entry after the definition was changed slightly. Compounds now come first, followed by self-explanatory syntactic examples and then idioms. Idioms are now given in bold, emphasising their status as sub-lemmas of a sort. Valency information – now somewhat more transparent – is found at the end of each numbered sense. Furthermore, although most examples in older editions of SLD are good, many infinitival phrases have been replaced by full sentences. In a learner's dictionary mainly intended for reception purposes, examples are extremely important. They must be self-explanatory, contain no difficult words and "[I]deally, they should evoke a little scene from everyday life, with prototypical actants" (Hult et al. 2010: 807). In the entry *väntar* ("wait(s)"), for example, instead of the infinitival phrase *vänta på bussen* ("wait for the bus"), the full sentence *de fick vänta på bussen i tio minuter* ("they had to wait for the bus for ten minutes") is given. In addition, all words in the dictionary are clickable.

The changes accomplished in the updating project are consistent with the conclusions of Pálfi & Tarp (2009) on the subject of learner's dictionaries. For instance, they recommend more synonyms and antonyms, more explicit valency information and lemmatisation of idioms. In his specific theories concerning learner's dictionaries, Tarp (2009: 199-200) also emphasises the importance of describing semantic relationships. This is realised in SLD through the copious links from the dictionary entries to the picture themes and, for verbs, the animations.

Let us now turn to the comment field data of the two web questionnaire surveys.

3 Types of Comments Expressed in the Two Sets of Comment Field Data

In the 2007 study, almost one third of the respondents, 110 out of 360, wrote something in the questionnaire's comment field. Interestingly, considering that the dictionary's target group is (relatively) recently arrived immigrants, almost 60% of respondents reported Swedish as their native language.

In terms of length, the shortest comment consists of one word and the longest of 130 words, with a median value of 14 words. In Table 1 the comments have been categorised into five types, each accompanied by a translated example. Comments belonging to more than one category were divided and placed into all appropriate types, which explains how the 110 comments in total turned into 152 parts.

Type of comment	Frequency		Examples (my translations)
	Absolute	Relative	
Criticism	53	35	*You only have easy words like "cat" and "dog", but people know those words…I need explanations for more difficult words.*
Suggestions for improvement	48	31,5	*It would be fun if there were more technical terms or words related to certain subjects like biology, etc.*
Praise	30	20	*I think Lexin and [SLD] is a very good site, I use it often and it is the best look-up site I have ever found.*
Other	17	11	*I think all the items in question 8 are important to have full understanding in a context.*
Nonsense	4	2,5	*ROCK THE WORLD*
Total	**152**	**100**	

Table 1: The 2007 study: comment types, frequency and an example of each type, in descending order.

Comments containing criticism are nearly as numerous as comments giving suggestions for improvement. Combined, they represent 66.5% of the comments. One fifth of the comments contain praise. The "Other" category covers comments of a more neutral nature or comments that did not directly relate to SLD. With a few exceptions, these comments contributed very little substantive information. The nonsense comments speak for themselves; they are few in number and contain no information of interest.

Let us now turn to the 2011 study. Nearly half the respondents, 371 out of 802, wrote something in the questionnaire's comment field. Gratifyingly, a full 67% of the respondents belong to the dictionary's target group: people whose native language is not Swedish and who arrived in Sweden fairly recently (during the 21st century).

In terms of length, the shortest comment consists of one word and the longest of 168 words, with a median of 17 words. In Table 2 the comments have been categorised into six types, each accompanied by a translated example. One column was added, called "Temporary change noticed".[1] Again, comments belonging to more than one category were divided and placed in all appropriate types. In total, the 371 comments turned into 565 parts.

---

1 The same week the questionnaire was displayed on the SLD website, the settings had been manipulated. Five categories had been removed from the dictionary articles with a view to indicating which categories users most preferred. Users were informed of the ongoing survey and encouraged to change the settings back to default if they wished. This change was mentioned in the comment field in about 15 questionnaires.

Type of comment	Frequency		Example (my translations)
	Absolute	Relative	
Praise	239	42	*Free service. Simple translation. Good examples on how to use words.* (Originally written in English)
Criticism	123	22	*You don't have enough words in Lexin.*
Suggestions for improvement	115	20	*Add more compounds, please!*
Other	69	12	*Didn't know about the pictures and videos.*
Temporary change noticed	15	3	*It's a pity you took away the examples and verb forms.*
Nonsense	4	1	*Huh.*
Total	**565**	**100**	

Table 2: The 2011 study: comment types, frequency and an example of each type, in descending order.

Compared to the 2007 study, the figures are distributed somewhat differently. Here, praise is the most frequent comment type, followed by criticism and suggestions for improvement. The positive comments (praise) are almost exactly equal in number to the suggestions for improvement and critical comments put together. Moreover, in relative figures there are twice as many positive comments compared with the 2007 study.

Notably, many comments ended up in the "Other" category. These are often part of longer comments; for instance, some respondents write something about themselves or their background ("I have dyslexia"), others say what other dictionaries they use ("I use Tyda instead"), and still others simply add a comment like "examples are important for immigrants" or "hope it remains free of charge". Compared with the "Other" comments in the 2007 study, these comments often contribute more or less substantive information.

The data reported so far indicate that respondents are happier with the revised version of SLD than with its unrevised counterpart. Comparing the figures from before and after the updating project, we see that praise has increased by 13.2 percentage points, whereas criticism has decreased by 12.4 percentage points. Comments in the "Other" category have declined by 0.8 percentage points. Let us now move on to see what information categories users comment on and what they say about them.

4   Information Categories Commented On

In the 2007 version of SLD, before the revision, a dictionary article could have up to nine information categories. These are mentioned 137 times in the comment field data of the 2007 study, distributed as shown in Table 3.

Information category	Positive context	Negative context	Other	Total	
				Absolute	Relative
1. Lemma	9	91	4	104	76
2. Pronunciation	1	7	4	12	9
3. Part of speech	1	-	1	2	1,5
4. Inflection	-	1	1	2	1,5
5. Meaning	-	8	-	8	6
6. Compound	1	1	1	3	2
7. Example	-	3	-	3	2
8. Phrase	-	2	-	2	1,5
9. Picture theme	-	1	-	1	0,5
Total	**12**	**114**	**11**	**137**	**100**

Table 3: Frequencies of SLD information categories in the 2007 study.

"Lemma" is by far the information category most frequently commented on (76%). "Pronunciation" and "Meaning" are in second and third place with 12% and 8% of the comments, respectively. The remaining categories are all rarely commented on. As shown, occurrences in a negative context are most common.

In the 2011 revised version of SLD, a dictionary article may have up to thirteen information categories. These are mentioned 332 times in the comment field data of the 2011 study, distributed as shown in Table 4.

Information category	Positive context	Negative context	Other	Total	
				Absolute	Relative
1. Lemma	36	126	10	172	51,5
2. Pronunciation	4	12	2	18	5
3. Part of speech	-	1	-	1	0
4. Inflection	5	11	1	17	5
5. Abbreviation	1	1	-	2	0,5
6. Meaning	8	13	2	23	7,5
7. Antonym	1	5	-	6	2
8. Compound	-	13	-	13	4
9. Example	9	30	3	42	13
10. Valency	-	4	1	5	1,5
11. Phrase	4	15	1	20	6
12. Picture theme	4	3	2	9	3
13. Video sequence	1	1	2	4	1
Total	**73**	**235**	**24**	**332**	**100**

Table 4: Frequencies of SLD information categories in the 2011 study.

Again, "Lemma" is by far most the most frequently mentioned category in the comments. "Example" is in second place, followed by "Phrases", "Pronunciation", "Inflection" and "Meaning". The least frequently mentioned are "Part of speech" and "Abbreviation". The remaining categories received between four and thirteen comments. Again, occurrences in a negative context are the most common, but are significantly fewer than in the 2007 study, with a difference of 12.4 percentage points. Thus, analogue to that indicated in the previous section, respondents seem happier with SLD after the revision than before.

Now let us look at more authentic examples from the comment field data. I will focus on what the respondents said about the information categories "Lemma", "Meaning" and "Example".

4.1 Voices on Lemma

The typical praise related to lemma simply states that the respondent often finds the words they searched for in SLD, and are satisfied with what can be found. Practically all the critical comments concerning lemma in both studies complain about the dearth of them. Many respondents are critical of the absence of more difficult words. Here are a few examples concerning lemma taken from the two sets of data (my translations):

Study 2007

(1) Some words you search for aren't there, even though they are real words.

(2) Quite a few psychological or political terms/words aren't there.

(3) It is bad that there are no compounds or particle verbs, which are what we immigrants need the most help with.

(4) I want the difficult words; the words included are mostly for beginners, or basic level.

Study 2011

(5) Many compounds are missing as well as special terms like *phonology* and *autism*.

(6) Hard to find the meaning of long words.

(7) Have a feeling only the most basic words are there, not other words.

(8) Sometimes important words are missing.

Approximately 3,000 new lemmas were added in the revision, including many particle verbs, and many synonyms and antonyms were added to the dictionary articles. In this respect, we have fulfilled the respondents' requests for more lemmas of different kinds. The figures speak for themselves: in 2007, 87.5% of the comments on lemma were negative, compared with 73% in 2011, while 8.7% of comments were positive in 2007 compared with 21% in 2011. Respondents are clearly less negative and more positive after the updating project. They remain, however, quite critical of the lemma selection and express more or less the same type of criticism in 2011 as in 2007. Indisputably, no dictionary will

ever succeed in fully satisfying users' need for more lemmas. A simple comment like "all words aren't there" is very telling in respect to users' expectations of dictionary's lemma selection. Along with the comments quoted above, this also reveals that many respondents do not clearly understand what type of dictionary the SLD is.

4.2 Voices on Meaning

In the 2007 study, there were no positive or neutral comments and only eight negative ones about the information category "Meaning". In the 2011 study, there are eight positive comments, thirteen negative comments and two neutral ones. Here are a few examples (my translations):

Study 2007

(9) More examples of the meaning of the word and how you can use the word.

(10) A definition of *naturopathic practitioner* is missing and an understandable explanation of *recruit*.

(11) Hard to find explanations of words.

(12) You only have easy words like *cat* and *dog*, but people know those words. I need explanations for more difficult words.

Study 2011

(13) Would like to have more examples of possible useful prepositions with a word and what meaning you get if you use a particular preposition.

(14) It should be possible to search for meaning/translation of expressions, not only single words.

(15) Can you try to give the exact meaning of words or an exact synonym?

(16) The definitions are good, short and concise and easy to understand.

The update included a thorough revision subdividing the senses of each lemma. Moreover, a great deal of effort was dedicated to more explicitly describing particle verbs which, as mentioned, often cause difficulties for learners of Swedish. Again, users' expectations are high, as they should be. In some respects, the respondents' views have been taken into account, for example in (9) and (13), and in other respects they have not, such as in (10). In a perfect world, all users would agree with the respondent quoted in (16).

4.3 Voices on Example

There were only three comments concerning "Example" in the 2007 study, all negative. Respondents want more examples, as expressed in the following comment: "have at least 2-3 examples for every word so you understand 100%". In the 2011 study, there are nine positive comments, thirty negative ones and three of a neutral nature. Here are a few examples (my translations):

(17) Too few examples on how to use words, hard to find synonyms, antonyms and proverbs.

(18) You should have more phrases and maybe a few more examples of each word or preposition.

(19) Good examples where you often find exactly the inflection you are looking for.

(20) Better if you could add more examples, particle verbs, past participle, proverbs, slang, which you
have to learn and use every day.

As thE senses of each lemma were subdivided, so were the examples, placing them in the immediate vicinity of the sense being illustrated. The update also added many morphological examples in terms of transparent compounds and several more, and more extensive, syntactical examples. In the updated version there is at least one example per lemma. This may not be enough in the opinion of the respondent quoted above, but examples are nonetheless much more numerous than was the case before the revision. What this respondent, and hopefully many others, acknowledges and appreciates is that the examples have been expanded to full sentences to more clearly illustrate the meaning of the lemma.

5   Conclusion

Taken as a whole, the users' comments on SLD can offer lexicographers valuable information about the needs of the dictionary's target group and might provide them with relatively concrete ideas on how to improve the dictionary. The question, however, was whether the respondents are happier with the SLD now than before the updating project. Comment field data from a web questionnaire survey before the update were compared with corresponding data from after the update. And, yes, respondents appear to be happier with the SLD now than before the revision. Firstly, there are significantly more positive comments after the revision than before. Secondly, the information categories commented on are more often set in positive contexts in the latter data. We also know that the majority of the respondents in 2011 belong to the dictionary's target group, in contrast to 2007. This suggests that the users who are happier with the SLD are also more representative of the intended users of the dictionary.

Even though respondents seem more satisfied with the updated version of SLD, the two sets of comment field data contain largely the same kind of criticism. Not unusually, many respondents have both good and bad things to say about the SLD. These circumstances indicate that one cannot have too much of a good thing. Or, as one respondent simply writes: "I think you can do it better, but it's not bad". Moreover, "Lemma" was the information category most frequently commented on, while the other categories were much less frequently commented on. Presumably, there is more information to be extracted from the comment field data.

It should be noted that the two sets of data were not fully comparable. The comments in 2007 add up to not quite one third of the number of comments in 2011. The first questionnaire contained ten questions and the second was extended to twenty questions. In addition, the texts preceding the comment field box were not identically worded.

6 Discussion

There are comments where respondents more or less explicitly demonstrate their awareness of what type of dictionary they are presently using, but there are also comments which clearly reveal the respondents' lack of understanding of what type of dictionary SLD is. This may not be a problem, since the dictionary is free and users are also free to go elsewhere if they are not satisfied with what they find. On the other hand, perhaps the purpose of SLD and all the Lexin dictionaries could be made more explicit on the website and on the internet as a whole. If not, at least the pictures and animations should be marketed since "for non-native speakers of the language, definitions, however skilfully written, are not usually the best way to convey meaning" (Lew, in press), and pictures and animations can concretise the meaning of a word in a very enlightening way. Surprisingly and somewhat disappointingly neither pictures nor animations are mentioned more than a handful of times in the comment field data. Admittedly, this does not necessarily mean that the respondents are unaware of these features, but unfortunately that might just be the case.

In the practice of dictionary making, professional lexicographers are undoubtedly the linguistic experts and users' potential contributions in this area may be limited. I would argue, however, that there are other equally important areas where users may very well have useful suggestions for improvement, particularly in relation to web-related issues like the user interface.

7 References

Hult, A.-K. (2008a). Användarna bakom loggfilerna. Redovisning av en webbenkät i Lexin online Svenska ord. In: *LexicoNordica*, 15, pp. 73-91.

Hult, A.-K. (2012). Old and New Study Methods Combined. Linking Web Questionnaire with Log Files from the Swedish Lexin Dictionary. In: R. Vatvedt Fjeld, J. M. Torjusen (eds.) *Proceedings of the 15th Euralex International Congress, Oslo, 7-11 August 2012*, pp. 922-928.

Hult, A.-K., Malmgren, S.-G., Sköldberg, E. (2010). Lexin – a Report from a Recycling Lexicographic Project of the North. In: A. Dykstra, T. Schoonheim (eds.) *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6-10 July 2010*, pp. 800-809.

Gellerstam, M. (1999): LEXIN – lexikon för invandrare. In: *LexicoNordica 6*, pp. 3-17.

Malmgren, S.-G. (2012). Från Svenska ord (Lexin) 3 till Svenska ord 4. In: B. Eaker, L. Larsson and A. Mattisson (eds.) *Nordiska studier i lexikografi 11. Rapport från Konferensen om lexikografi i Norden, Lund 24-27 May 2011*, pp. 454-465.

Lew, Robert in press. User-Generated Content (UGC) in Online English Dictionaries. In: A. Klosa, A. Abel (eds.) *OPAL – Online publizierte Arbeiten zur Linguistik* 2013. (Preprint).

Lexin Online, Statistics. Accessed at: http://lexin2.nada.kth.se/statistik/html [29/03/2014].

Pálfi, L.-L., S. Tarp (2009). Lernerlexikographie in Skandinavien – Entwicklung, Kritik und Vorschläge. In: *Lexicographica. International Annual for Lexicography 25/2009*. Tübingen, pp. 135-154.

Tarp, S. (2008). Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner´s Lexicography. Tübingen: Max Niemeyer.

Swedish Lexin Dictionary Online. Accessed at: http://lexin2.nada.kth.se/lexin [29/03/2014].