

The Ideal Number of Lemmas in an Ideal Accounting Dictionary

Abstract

Lemma lacunas in dictionaries are a traditional focus area for lexicographers, but the opposite problem, which we choose to call **lemma flooding**, has received very little attention. The study of this flooding could be relevant in order to save lexicographers spending thousands of hours producing dictionary entries which nobody reads.

In Bergenholtz/Norddahl (2012) we showed that during a three-year period less than 33% of all dictionary articles out of 18 million dictionary consultations were consulted in a dictionary with 111,000 entries. We examined nine possible reasons why a given word might not be of interest to users and consequently could be ignored in order to avoid lemma flooding. We tried to demonstrate that while it is not possible to completely avoid lemma flooding, implementing a relatively simple rule could minimize it. But in reality the results were quite disappointing, because there were no clear rules or methods to avoid lemma flooding.

Now we will try the same kind of analysis of log files for the English-Danish and the Danish-English Accounting Dictionaries. We see here that there are differences between different dictionaries (monolingual for English and Danish and bilingual for English-Danish and Danish-English). We will try to give some explanations, but must admit beforehand that we have not found satisfying explanations which could lead to a plan for future accounting dictionaries or other economic dictionaries thus avoiding the production of never used dictionary articles.

1. Introduction

If dictionary users look up a word in a dictionary and very rarely find what they are looking for, they might stop using the dictionary. Critics of a dictionary very often focus on such problems. Therefore, it is no surprise that dictionary authors and researchers focus on how to avoid having any lemma lacunas.

However, very little attention has been given to unnecessary lemmas in dictionaries. From Bergenholtz/Norddahl (2012) we know that more than 66% of all articles in *Den Danske Netordbog* have never been consulted over a 3-year period. More precisely, we are speaking about 37,238 dictionary articles looked for, but also 74,254 dictionary articles not looked for a single time in 18 million lookups. It is expensive and time consuming to produce unnecessary lemmas which could have been used on more relevant tasks.

The lack of relevant lemmas in a dictionary is called lemma lacunas. But we do not have an appropriate word for the many unnecessary lemmas in dictionaries. We propose to use the term **lemma flooding** (Danish **Lemmaoverflow**, German **Lemmaüberfluß**).

For printed dictionaries lemma flooding is a problem for the user because it makes the dictionary big, heavy and more time consuming to consult. The ideal dictionary only contains the information which the user needs and nothing else. In an electronic dictionary, however, it is not a disadvantage for the user that the dictionary contains lemmas which the user does not need. It is only a problem for the lexicographer.

Bergenholtz/Johnsen (2005) are sceptical about whether we can ever find any systematical descriptions of never used dictionary articles:

* Henning Bergenholtz
Centre for Lexicography
Department of Business Communication
School of Business and Social Sciences – Aarhus University
Jens Chr. Skous Vej 4
DK-8000 Aarhus C
hb@bcom.au.dk

* Bjarni Norddahl
Ordbogen A/S
Billedskærervej 8
5230 Odense M
bn@gigabit.dk

Furthermore, it is not possible to discern a distinct pattern on the basis of this examination, e.g. that certain types of words, such as semantic or orthographic variants, are never requested. A systematic description of the requested words compared with the non-requested words is thus not possible, and it remains unclear whether such an investigation would be of practical use to lexicographers. (Bergenholtz/Johnsen 2005: 139f)

The results of Bergenholtz/Norrdahl (2012) are similarly disappointing without any really good rules to avoid lemma flooding. Both studies focus on general language dictionaries. In this article we will focus on specialized accounting dictionaries to search for a pattern and see how big the lemma flooding problem is compared to general language dictionaries.

2. Explanations for lemma flooding

In Bergenholtz/Norrdahl (2012) we have put forward nine explanations for lemma flooding in general language dictionaries:

1. No system

The possibility that there is no explanation for why some lemmas are looked up, while others are not (Bergenholtz/Johnsen 2005). This is also a possibility in accounting dictionaries.

2. Not relevant words

This category is especially relevant for specialized dictionaries including accounting dictionaries. If an accounting dictionary includes words which would be more suitable in a medical dictionary, this would be considered as lemma flooding. We did not find such words in any of the accounting dictionaries that we analysed and therefore cannot explain any lemma flooding occurring in these dictionaries.

3. Words known by everybody

If the user already knows the answer he will never search for it. This could be true if everyone using accounting dictionaries was an expert. But students and others who do not necessarily know even basic accounting terminology also use accounting dictionaries.

4. Easily understandable composite

In the Danish language easily understandable composites occur very frequently. Since we found no such system in *Den Danske Netordbog*, it is unlikely that this should be the case in English.

5. Foreign words

We see the same search behaviours with foreign words. Which foreign words are searched for and which are not seems to be random.

6. Neologism

New words in the language are among the lemmas most searched for. They seem to never be among the zero look-ups. This is more a category to avoid lemma lacunas than lemma flooding.

7. Words not used anymore

Some words almost disappear from a language over time and this might be a reason for why they are not looked up. If the words are not used anymore, this could explain why they are not searched for. The accounting dictionaries used for this research do not contain words that are not used anymore.

8. Words from specialized fields

In a general language dictionary the user might not expect to be able to find lemmas from a specialized field and will therefore go directly to a more relevant dictionary. This category

is not relevant for accounting dictionaries, since the lemmas are from a specialized field and should always primarily contain words from the field of accounting.

9. Infrequent words

In order to produce dictionaries lexicographers often use frequency, which is a bad criteria. From our data it is also clear that this is not a criteria which can be used for specialized dictionaries either, since relatively common words such as “accepted”, “appendix” and “deficiency” have zero look-ups whereas a much more infrequent word such as “amortisation” has 88 look-ups.

We had the clear assumption that a much bigger part of the lemma stock would be relevant for the user with a specialized dictionary compared to a general language dictionary. But also that there might be a difference between monolingual and bilingual dictionaries.

3. Results

We have looked in the log files of the *English Accounting Dictionary*, the *Danish Accounting Dictionary*, the *English-Danish Accounting Dictionary* and the *Danish-English Accounting Dictionary* at Ordbogen.com. The results are compared with the results of Bergenholtz/Norrdahl (2012). In the article of *Den Danske Netordbog* we did not include the total number of look-ups and therefore it is not included in the lists below. As the total number of lemmas increases constantly in all the dictionaries, the number of entries is that of the date that we looked in the log files for the specific dictionaries, and the number is even higher today.

	<i>English Accounting dictionary</i>	<i>Den Danske Netordbog</i>
Number of lemmas	6,914	111,492
Lemmas looked up	2,795	37,238
Lemmas not looked up	4,119	74,254
Hit percentage	40.43%	33.40%
Number of look-ups	37,902	-
No. of searches with a result	12,849	17,893,973

Table 1. *English Accounting Dictionary* versus *Den Danske Netordbog*

As we can see, the results are only slightly better than the general Danish monolingual dictionary. 59.57% of the lemmas have never been consulted which is quite disappointing for the lexicographers. However, the audience is primarily Danish and they might make more spelling errors which can explain some of the disappointing results.

In a search for a system as to which lemmas the users search for and which they do not we have generated a list of the 10 most searched for articles and 10 randomly selected articles with zero look-ups:

Rank	Lemma	Look-ups
1	underwriting	214
2	balance	148
3	accounting	144
4	insolvent	106
5	revenue	103
6	deemed cost	94
7	amortisation	88
8	provision	83
9	asset	81
10	account	74

Table 2. The 10 most searched for articles in the *English Accounting Dictionary*

These are all quite normal words and it is therefore no surprise that they are looked up. It also underlines that the third possible explanation (words known by everyone) cannot be true for accounting dictionaries.

To see which lemmas the lexicographer could have avoided producing we need to take a look at which lemmas have zero look-ups:

Lemma	Look-ups
accessory	0
accepted	0
appendix	0
deficiency	0
degressive	0
fragmented	0
franchisee	0
invest	0
leased	0
liable	0

Table 3. Randomly selected lemmas with zero look-ups in the *English Accounting Dictionary*

Again, we see very common words in the list. There seems to be no logic distinction between the list with most look-ups and the list with zero look-ups. From our previous explanations it seems that only the “no explanation” option is true.

Let us see how this compares to the *Danish Accounting Dictionary*:

	<i>Danish Accounting Dictionary</i>	<i>Den Danske Netordbog</i>
Number of lemmas	7,345	111,492
Lemmas looked up	4,713	37,238
Lemmas not looked up	2,632	74,254
Hit percentage	64.17%	33.40%
Number of look-ups	90,158	-
No. of searches with a result	53,091	17,893,973

Table 4. *The Danish Accounting Dictionary* versus *Den Danske Netordbog*

This time, we see that the hit percentage has increased to 64.17 % which is a lot higher than both the *English Accounting Dictionary* (40.43%) and *Den Danske Netordbog* (33.40%). One explana-

tion for the difference could be that the audience is Danish. Also, it seems that people more often find what they are looking for in the first place.

Rank	Lemma	Look-ups
1	børs (Engl. exchange)	777
2	årsregnskab (IFRS Engl. financial statements)	557
3	regnskab (IFRS Engl. financial statements)	540
4	nedskrivning (Engl. 1. write-down, 2. UK Engl. impairment loss)	380
5	tilgodehavende (Engl. receivable)	273
6	afskrivning (Engl. 1. amortisation , 2. Engl. write-off, 3. IFRS Engl. amortisation charge)	268
7	omsætning (Engl. IFRS revenue)	265
8	egenkapital (Engl. equity)	236
9	moms (Engl. VAT)	202
10	balance (Engl. balance sheet)	196

Table 5. The 10 most searched for articles in the *Danish Accounting Dictionary*

In the above list we see the 10 most searched for articles in the *Danish Accounting Dictionary* which look somewhat similar to the *English Accounting Dictionary*.

Lemma	Look-ups
aconto-beløb (Engl. amount paid on account)	0
afgiftspligt (Engl. tax liability)	0
analysere (Engl. review)	0
egenkontrol (Engl. self-review)	0
erstatningsbeløb (Engl. amount of damages)	0
financiel (Engl. financial)	0
finanschef (Engl. treasurer)	0
importafgift (Engl. import duty)	0
inflationsrate (Engl. inflation rate)	0
momsperiode (Engl. VAT period)	0

Table 6. Randomly selected lemmas with zero look-ups in *The Danish Accounting Dictionary*

Just like *The English Accounting Dictionary* there is no clear pattern as to which lemmas the users search for and which they do not.

This much for the monolingual accounting dictionaries. Now we will take a first look at bilingual dictionaries. We start with the *English-Danish Accounting Dictionary*:

	<i>English-Danish Accounting Dictionary</i>	<i>Den Danske Netordbog</i>
Number of lemmas	6,914	111,492
Lemmas looked up	4,551	37,238
Lemmas not looked up	2,363	74,254
Hit percentage	65.82%	33.40%
Number of look-ups	104,110	-
No. of searches with a result	45,984	17,893,973

Table 7. *English-Danish Accounting Dictionary* versus *Den Danske Netordbog*

There is a slight increase in the hit percentage but it is almost the same as the monolingual Danish accounting dictionary. Just like the *English Accounting Dictionary* the hit percentage might be influenced by the fact that we have a Danish audience. This can also be seen in the number of searches with a result which is under 50% of the total number of look-ups.

Rank	Lemma	Look-ups
1	revenue	343
2	accounting	237
3	underwriting	234
4	impairment	233
5	equity	225
6	balance	204
7	provision	198
8	account	176
9	depreciation	164
10	cash	154

Table 8. The 10 most searched for articles in the *English-Danish Accounting Dictionary*

When looking at the list of most searched for lemmas we can see that there are many similarities between the monolingual English accounting dictionary and the bilingual English-Danish accounting dictionary. It does not, however, contribute to finding any good explanation for how to avoid lemma flooding.

Lemma	Look-ups
accessory	0
amortizable	0
annually	0
appendix	0
bookkeeper	0
borrow	0
client	0
committee	0
complain	0
incumbrance	0

Table 9. Randomly selected lemmas with zero look-ups in the *English-Danish Accounting Dictionary*

Randomly selected zero look-ups do not give any suggestions for explaining the lemma flooding in this dictionary either.

The last dictionary we will look at is the *Danish-English Accounting Dictionary*:

	<i>Danish-English Accounting Dictionary</i>	<i>Den Danske Netordbog</i>
Number of lemmas	7,345	111,492
Lemmas looked up	5,679	37,238
Lemmas not looked up	1,666	74,254
Hit percentage	77.32%	33.40%
Number of look-ups	274,005	-
No. of searches with a result	165,983	17,893,973

Table 10. *Danish-English Accounting Dictionary* versus *Den Danske Netordbog*

77.32% is the highest hit percentage which we have seen to date. It is more than twice as good as that of *Den Danske Netordbog*. This is a much more satisfying number for the lexicographer but we do still have a relatively high number of lemmas which have never been used. The Danish-English language pair combination is also by far the most used combination of all the accounting dictionaries

Rank	Lemma	Look-ups
1	årsregnskab (IFRS Engl. financial statements)	1,530
2	drift (Engl. operation)	1,456
3	afskrivning (IFRS Engl. amortisation)	1,417
4	nedskrivning (Engl. 1.write-down, 2. UK Engl. impairment loss)	1,317
5	tilgodehavende (Engl. receivable)	979
6	regnskab (IFRS Engl. financial statements)	912
7	børs (Engl. exchange)	799
8	kapitalandel (IFRS Engl. equity investment)	768
9	omsætning (IFRS Engl. revenue)	719
10	egenkapital (equity)	673

Table 11. The 10 most searched for articles in the *Danish-English Accounting Dictionary*

Apart from a different order in the rank we can see that 8 out of 10 lemmas in the monolingual Danish accounting dictionary are identical with the Danish-English Accounting Dictionary. This suggests that we have the same lack of logical explanation as before.

Lemma	Look-ups
balanceskema (Engl. balance sheet format)	0
brugsfordel (Engl. use benefit)	0
bruttooverskud (Engl. gross profit)	0
certifikat (Engl. certificate)	0
dokument (Engl. document)	0
ejerprincip (Engl. proprietary view)	0
erklæringsdato (Engl. date of report)	0
financiere (Engl. finance)	0
fraskrivelse (Engl. waiver)	0
genleje (Engl. re-lease)	0

Table 12. Randomly selected lemmas with zero look-ups in the *Danish-English Accounting Dictionary*

And as is the case with the other dictionaries, we still cannot find a good explanation by looking at the list of zero look-ups.

4. Conclusion

The ideal dictionary gives the user the needed information and nothing else. So the ideal number of lemmas in an accounting dictionary would be to have no lemma lacunas and no lemma flooding. Traditionally, most lexicographers have only lemma lacunas in mind, but not the vast amount of wasted time and effort produced on lemma flooding, which in some cases may consist of more than 50% of all lemmas.

Is it possible to avoid lemma flooding? No! Just like it is not possible to avoid lemma lacunas. Nothing from our research indicates that lemma flooding can be avoided. Neither does it seem that we can currently set up any good rules in order to avoid lemma flooding.

One way to handle the issue was used in *Ordbogen.com's Danish-English* general language dictionary where in the beginning some degree of lemma lacunas was accepted during a beta pe-

riod. After that, log files were analysed daily and missing lemmas were added to the dictionary. Another way is used in the accounting dictionaries where data are added when users contact us.

5. Bibliography

5.1. Dictionaries

Danish Accounting Dictionary = Sandro Nielsen, Lise Mourier and Henning Bergenholtz 2014: *Den Danske Regnskabsordbog: Betydning*. Database: Richard Almind og Jesper Skovgård Nielsen. Odense: Ordbogen.com. (Accessed January 2014)

Danish-English accounting dictionary = Sandro Nielsen, Lise Mourier and Henning Bergenholtz 2014: *Den Danske Regnskabsordbog: Betydning*. Database: Richard Almind og Jesper Skovgård Nielsen. Odense: Ordbogen.com. (Accessed January 2014)

Den Danske Netordbog = Henning Bergenholtz in cooperation with Heidi Agerbo and with the participation of Andreas Bodilsen, Sofie Weiss Dohrn, Kathrine Brosbøl Eriksen, Jakob Esmann, Helene R. Gudmann, Christian Kjølhede, Jane Nguyen, Mia Lybkær Kronborg Nielsen and Henrik Thers. Database: Richard Almind og Martin Carlsen 2014: *Den Danske Netordbog*. Odense: Ordbogen.com.

English accounting dictionary = Sandro Nielsen, Lise Mourier and Henning Bergenholtz 2014: *Den Danske Regnskabsordbog: Betydning*. Database: Richard Almind og Jesper Skovgård Nielsen. Odense: Ordbogen.com. (Accessed January 2014)

English accounting dictionary = Sandro Nielsen, Lise Mourier and Henning Bergenholtz 2014: *Den Danske Regnskabsordbog: Betydning*. Database: Richard Almind og Jesper Skovgård Nielsen. Odense: Ordbogen.com. (Accessed January 2014)

English-Danish Accounting dictionary = Sandro Nielsen, Lise Mourier and Henning Bergenholtz 2014: *Den Dansk-Engelske Regnskabsordbog: Oversættelse*. Database: Richard Almind og Jesper Skovgård Nielsen. Odense: Ordbogen.com. (Accessed January 2014)

Ordbogen.com Dansk-Engelsk = Bjarni Norddahl and Michael Walther 2006-2014: *Ordbogen.com Dansk-Engelsk*. Odense: Ordbogen.com.

5.2. Other literature

Bergenholtz/Norrdahl 2012 = Henning Bergenholtz and Bjarni Norrdahl 2012: Ordbogsartikler, som ingen læser. *LexicoNordica* 19, 207-223.

Bergenholtz/Johnsen 2005 = Henning Bergenholtz and Mia Johnsen 2005: Log Files as a Tool for Improving Internet Dictionaries. *Hermes* 34, 117-141.