

Anglicisms and Language-Internal Neologisms: Dealing with New Words and Expressions in

The Danish Dictionary

Lars Trap-Jensen

Society for Danish Language and Literature

ltj@dsl.dk

Abstract

The Danish Dictionary, a corpus-based online dictionary, contains just over 100,000 entries. The dictionary is updated on a regular basis, with new versions published two or three times a year. Whenever an update is released, it almost always becomes the object of public attention. The media love new words and usually assume that a new word in the dictionary is also a new word in the language – a neologism. Of course, popular belief is far from the truth: many newly published words have been in the language for a long time, but were perhaps too infrequent to be included previously.

Given their popularity, neologisms are obviously interesting for the dictionary staff, and in this paper I will analyse those that have been included recently, and consider whether special selection criteria should apply. The editors do not use a specific method to detect neologisms in particular, but we have various tools to assist us in finding lemma candidates in general, and we can also analyse the updates that have been published in recent years. I will pursue both these approaches, addressing questions such as the following:

- (1) What broad types of neologisms exist and what are their characteristics?
- (2) How does pressure from English affect the vocabulary of the dictionary?
- (3) Are Anglicisms dominant or used increasingly over time as compared with language-internal neologisms? Does globalisation promote the import of words from other languages, too?
- (4) Do dictionary users suggest and look up neologisms, and in particular Anglicisms, more often than other words?

Although the notion ‘neologism’ pertains to a range of linguistic phenomena, I will in this context confine myself to words and multi-word expressions as (potential) entries.

Keywords

corpus-based lexicography, lemma selection criteria, Anglicisms, dictionary use, neologisms

Introduction

Linguistically, a neologism can be regarded as any new linguistic phenomenon observable in a language, including a new lexical item, whether on word, sense or phraseological level, or a change in pronunciation, morphology, grammatical feature or usage pattern. However, in this context, only lexical items will be considered and, in particular, new words and multi-word expressions.

According to most definitions, it is central to the meaning of a neologism that it is a newly coined word. Neologisms are sometimes further defined as being not yet fully established within the entire language community¹. If they are only marginally established, one would not expect them to be included in a general-purpose dictionary.

However, in this context, I am not particularly concerned with theoretical definitions of the term. I will analyze and account for the way neologisms are dealt with in one particular dictionary, *The Danish Dictionary*, and in this context, what constitutes a neologism is something which is entirely determined by the lexicographic practice of that dictionary.

New Words and Neologisms

Seen from the perspective of the dictionary, a new word or expression is simply one that has not previously been included in the dictionary. Not all new words and expressions are neologisms – far from it. Many words added to the dictionary have been in the language for a long time, but the

1 cf. Bußmann (1990) *Neologismus*: Neugebildeter sprachlicher Ausdruck (Wort oder Wendung), der zumindest von einer Teil der Sprachgemeinschaft, wenn nicht im allgemeinen, als bekannt empfunden wird (‘Newly formed linguistic expression (word or phrase) that is known by at least a part of the linguistic community, if not in general’).

editors did not include them previously, perhaps because the words were not frequent enough, or because of insufficient resources, or for whatever practical reason. However, a subset of new words are also new in the language, and they are the ones that will be treated as neologisms here. One question that needs to be answered before we proceed to the analysis is this: how new must a new word be in order to count as a neologism? There does not seem to be any generally accepted period of time for which a word has to have been in use before it ceases to be considered new², and in our case, we have used, somewhat arbitrarily, the year 1990 as the cut-off point, mostly for practical, project-internal reasons. Nevertheless, here, I consider neologisms to be words or multi-word expressions in the dictionary for which the earliest recorded use is 1990 or later.

The Dictionary

The Danish Dictionary (*Den Danske Ordbog*, henceforth DDO) is a corpus-based monolingual dictionary of Danish describing the modern language from 1950 onwards. It was first published as a paper dictionary in six volumes, 2002-2005, and it was then converted into an e-dictionary. It has been online since 2009, available to the public free of charge and without advertisements, with regular updates two or three times a year. The dictionary is sponsored jointly by the Danish Ministry of Culture and a commercial organization, the Carlsberg Foundation.

Neologisms in DDO

At the time of writing (April 2019), the dictionary contains just over 100,000 entries, of which 10,000 have been added to the dictionary within the last three years. This new material forms the

² cf. usage note in the Wiktionary entry *neologism*: “There is no precise moment when a word stops being ‘new’, but 15–20 years is a common cutoff (corresponding to one generation growing up potentially familiar with the word, depending on how common it is). Acceptance of a word as valid by dictionaries or by a significant portion of the population is sometimes mentioned as additional conditions. Some neologisms become widespread and standard (such as new chemical element names), others remain rare or slangy.”

object of investigation in this paper. Table 1 sums up the distribution of words, multi-word expressions and senses, and for each category the proportion of neologisms is given.

Table 1

	Added 2016-2019, total	of which neologisms
New words	10,207	806 ³
New MWEs	51	21
New senses	149	21

MATERIAL ADDED TO DDO 2016-2019

These figures do not begin to cover all the neologisms that have been created since 1990. In total, DDO contains 2,986 neologisms from this period. By comparison, *Nye Ord i Dansk*, a specialized dictionary of new Danish words, registered about 4,000 words, multi-word expressions and senses in the same period, and these also represent just a selection. Nevertheless, the entries can give us an idea of which parts of the vocabulary are particularly productive and what word-formation patterns are used.

Neologisms and Domains

As all senses in the dictionary are assigned to a subject area whenever possible (which is not necessarily visible to the user, and is not to be confused with domain markers proper), it is possible to see which domains give rise to new lexical items. In Table 2, domains which occur in more than ten entries have been ranked. While computing, sports, food and communication are the most productive domains, a broad range of domains contributes to vocabulary growth. The cut-off point

³ Potentially, another 317 entries could belong to the group of neologisms, but these await confirmation of earliest record in an external archive to which the editors do not have immediate access. How many of the entries will be included in the dictionary is uncertain, and therefore only the 806 already confirmed will be included in this paper.

is set here – for practical reasons – to ten entries, but the 30 domains listed in Table 2 do not even represent the median: another 69 domains are represented in the lower-ranking domains.

Table 2

Domain	no. of articles	Domain	no. of articles	Domain	no. of articles
computing	105	society	26	family	15
sport	68	music	25	commerce	14
food	53	leisure	23	clothes	13
communication	51	television	21	religion	13
business	44	crime	20	home	12
sex	35	press	18	technology	11
politics	34	culture	17	games	10
medicine	33	gastronomy	17	social care	10
economy	28	drinks	17	literature	10
psychology	28	education	16	film	10

DISTRIBUTION OF NEOLOGISMS ACROSS DOMAINS

It seems safe to conclude that lexical innovation takes place across a wide range of areas, with computing, sports, food and communication as the most productive domains.

Morphological Characteristics of Neologisms

A further question to ask is whether particular word-formation patterns can be identified for neologisms, and if so, which are the dominant ones. For the above set of 806 neologisms, the main categories of word formation are described below.

Compounds and derivations of Danish words. Language-internal compounds and derivations are based on existing Danish words or morphemes. In Danish and the other Nordic languages, as well

as in German and Dutch, compounding is an extremely common way of forming new words, and traditionally this – together with the second most productive method, derivation – is by far the most common way to create new linguistic material, whether new words for new phenomena: *akutlæge* (‘emergency doctor’), *duftlys* (‘scented candle’), *eurosamarbejde* (‘Euro co-operation’), *fedtforskrækket* (‘fat-scared’), *hektarstøtte* (‘per acre support’), *ældremad* (‘food for senior citizens’), or new words for old phenomena: *fastnettelefon* (‘landline telephone’), *partiskift* (‘(political) party change’). In the present material, language-internal compounds and derivations account for 52% of the neologisms.

Direct borrowings. Direct borrowings from other languages make up 27% of the neologisms: *crowdfunding*, *dukkah*, *emoji*, *lassi*, *streetfood*, *tiqui-taca* etc. In this group, we also find newly coined internationalisms such as *heteronormativ*, *bioterrorist* and *disambiguere*. The Danish language has always been fairly open to borrowings, but the source of import has of course varied from one era to the next, depending on commercial partners and dominant cultural influence. Over the last century, English has been the primary supplier of foreign words, and in the present material direct borrowings from English account for 21 out of the 27 percentage points of neologisms that are direct borrowings. The remaining 6% are distributed over a variety of languages, the most frequent of which are modern internationalisms based on Greek and Latin such as *antropocæn*, *autofiktion*, *autogyro*, *emerita*, *neurogastronomi* and *nocebo*. Food and drink is the most productive domain in this category. Examples are *caipirinha* (Portuguese), *dim sum* (Cantonese), *dukkah* (Arabic), *durum* (Turkish), *kimchi* (Korean), *kaki*, *kombucha*, *mirin*, *nigiri*, *ramen*, *soba*, *tempura*, *udon*, *yuzu* (Japanese), *lassi* (Hindi), *margarita*, *manchego* (Spanish), *panna cotte* (Italian). The rest is a mixture of individual words from various domains and languages such as *bong* (Thai), *emoji* (Japanese), *gruppetto* (Italian), *nada* (Spanish) and *panna* (Dutch). It is probably true that many of the more remote borrowings have not been imported directly into Danish, but have entered via other languages, notably English. Nevertheless, they bear witness to globalisation and the fact that we are

exposed to impressions from the remotest parts of the world, via mass media and through changing patterns of travel and leisure time.

Blends. Blends consist of a borrowing in combination with a Danish word or morpheme, most often in the form of compounds or derivations: *blogindlæg* (*blog* + Danish *indlæg* ‘contribution’), *brunche* (‘to have brunch’, *brunch* + Danish verbal suffix *-e*), *edamamebønne* (Japanese *edamame* + Danish *bønne* ‘bean’). Altogether, 16% of the neologisms belong in this group, and once again English is the dominant source for the foreign part, accounting for 13 out of the 16 percentage points of neologisms which are blends. Examples from other languages include *gojibær* (Chinese *goji* + Danish *bær* ‘berry’), *serranoskinke* (Spanish *serrano* + Danish *skinke* ‘ham’), *tandoorikylling* (Urdu *tandoori* + Danish *kylling* ‘chicken’), *salafisme* (Arabic *salaf* + *-isme*), *tantralsex* (Sanskrit *tantra* + *sex*), *teriyakisovs* (Japanese *teriyaki* + Danish *sovs* ‘sauce’).

Loan translations. On the expression plane, loan translations are not recognizable as borrowings, but their forms are motivated by foreign words which have been translated literally into Danish. Numerically, they are unusually infrequent in the present material (or I have not been aware of their foreign source), making up only 3% of the material⁴. As with the previous groups, English is the main source of influence: *elevatortale* (from *elevator speech*), *fagfællebedømmer* (based on *peer reviewer*: *fagfælle* ‘colleague’ + *bedømmer* ‘evaluator, reviewer’), *gademad* (based on *street food*: *gade* ‘street’ + *mad* ‘food’), *garagesalg* (based on *garage sale/car boot sale*), *honningfælde* (based on *honeytrap*: *honning* ‘honey’ + *fælde* ‘trap’), *højreklik* (based on *right click*: *højre* ‘right’), *kaffebordsbog* (based on *coffee-table book*). The only non-English example I have found is *babyluge* (based on German *Babyklappe*: *Klappe/luge* ‘hatch’).

⁴ Other investigations (cf. <https://dialekt.ku.dk/sproghistorie/engelskidansk>) have concluded that nine out of ten borrowings from English come in the form of loan translations and only 10% as direct loans.

Although not very frequent in the present material, this type is lexicographically interesting, and in terms of language policy it is important what position the dictionary takes: if a language-internal alternative – whether an independent coinage or a loan translation – exists, should the parallel direct borrowing be suppressed, should both words be included, or should only the more frequent of the two be entered? This question will be taken up below in connection with Anglicisms.

Other types of word formation. The remaining word formations in the material consist of various abbreviations and portmanteaus, with Danish, foreign and mixed origins. Examples of abbreviations are *DIY*, *HDMI*, *LGBT-*, *MDF*, *UMV* (Danish abbreviation for *undervisningsmiljøvurdering* ‘teaching environment evaluation’) and *VPN*, whereas portmanteau examples include *bygotto* (Danish *byg* ‘barley’ + part of Italian *risotto*), *blandinavisk* (Danish *blande* ‘mix, blend’ and part of *skandinavisk*, i.e. a mixed Scandinavian lingua franca), *glokal* (Danish version of *glocal* ‘global and local’) and *fleksitar* (same as English *flexitarian*). This word-formation method is not very productive, and this group is the smallest of the five comprising only 1–2% of the material.

Even so, it is remarkable that we find several portmanteaus in the material. This word-formation mechanism is otherwise rare in Danish and leads one to speculate whether the high prestige of English has boosted the number of words of this type, be they language-internal formations or in the form of borrowings or blends.

Anglicisms

In Denmark, as in many other countries, public opinion on Anglicisms is divided. Here as elsewhere, English is the first foreign language as well as the most prestigious one and as such it is the primary source of lexical borrowing in modern times. Borrowings from other languages are often imported as internationalisms through the English language, e.g. *emoji* (via English from Japanese), *fashionista* (English *fashion* + Spanish *-ista*) or *zika* (via English from the Zika forest in Uganda).

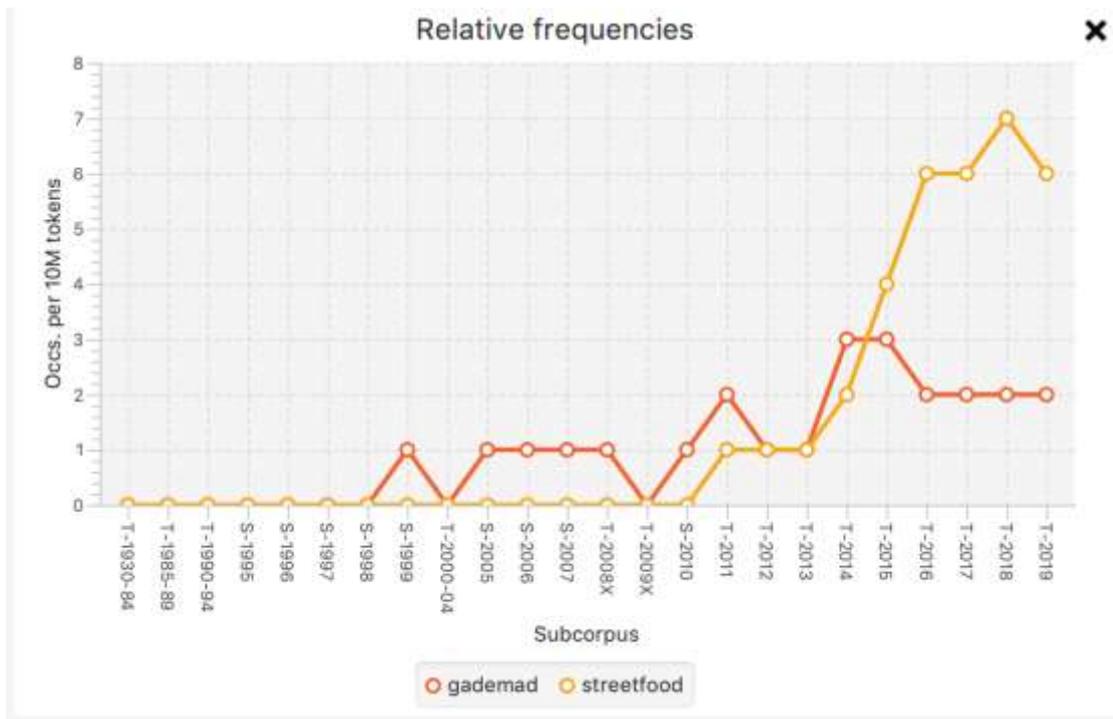
Having always been fairly open to lexical borrowings, the Danish language has in the past been heavily influenced by Middle German and French in particular, and only in the last century has English become the primary source of imported words. Being a Germanic language, English word formation is not dissimilar to Danish, and as Danish morphology is not highly complex, English words are not hard to absorb, often with an approximated pronunciation and at first with defective morphology until further assimilation takes place. In this way, Danish absorbs quite a large number of English borrowings – although some public debaters exaggerate the influence of English, which is nowhere near the Middle German impact on Danish in the Middle Ages. Recent studies suggest that Anglicisms make up no more than approximately 2% of running words in an average Danish text (cf. Gottlieb 2015).

In the present selection of 806 neologisms, a total of 188 Anglicisms were identified, including direct borrowings and loan translations, as well as blends with domestic words and morphemes. Before taking a closer look at them, let us first consider the lemma selection procedure. **Editorial policy on Anglicisms.** In order for an Anglicism to be included in DDO, it must – like all other words – fulfill the numerical requirements that corroborate its established usage in the common language. This is probably the easiest part of the process, but in addition, the lexicographer must consider whether the candidate word is an instance of borrowing or code-switching. In a situation where the majority of the population master the English language at a fairly advanced level, it is not unusual for people to switch more or less freely between them and insert words from English into an otherwise clear Danish context. If they have a clear understanding that two different languages are involved, the words used belong in two different dictionaries. But often it is hard to tell what the speaker's understanding is when only the text produced can be observed. In this case, the lexicographer must use his or her judgment to make a decision. Sometimes textual cues can help in the decision: does the word show signs of assimilation into the other language in terms of pronunciation or morphology? Is it used in a slightly different sense than in the source language?

An example of the former would be the Danish pronunciation [ˈpadəl, bɔːd] noted for *paddleboard* as opposed to the /ˈpædlbɔːd/ (here quoted from Oxford Learners' Dictionary) pronounced with different vowel qualities and a 'dark' l by native English speakers (although this is rarely represented in phonetic notation). Morphological examples are plenty: nouns must be assigned a grammatical gender in Danish, the plural may keep the English -s (as in *dramedy*, plural *dramedies*) or adapt to the Danish paradigm (as in *dongle*, plural *dongler*), or vary between them (as in *chatroom*, plural *chatroom* or *chatrooms*), and definite forms in the singular and plural may or may not be added. An example of different usage occurs when the English word *pride* is borrowed in only one of its senses, 'a parade where sexual minorities proudly demonstrate their culture and claim equal rights', or when *trail* in a Danish context is used only to denote running on tracks and paths in nature. In these cases, the original English word takes on one particular meaning as a borrowing and may develop even further from there. The fact that the English word *fuck* is used in Danish as a swearword too can hardly be interpreted as an instance of codeswitching, even if it maybe was when first used. Since its arrival into Danish, the word has developed connotations and usage patterns that are distinctly different from its usage in English. For this reason, it is natural to treat it as a Danish loanword from English which has a description of its own, distinct from what you find in English dictionaries.

Anglicisms vs. language-internal forms over time. In our corpus interface, we have developed a function that allows us to follow the frequency of words over time. In Figure 1 an example is given for the pair *gademad/streetfood*.

Figure 1

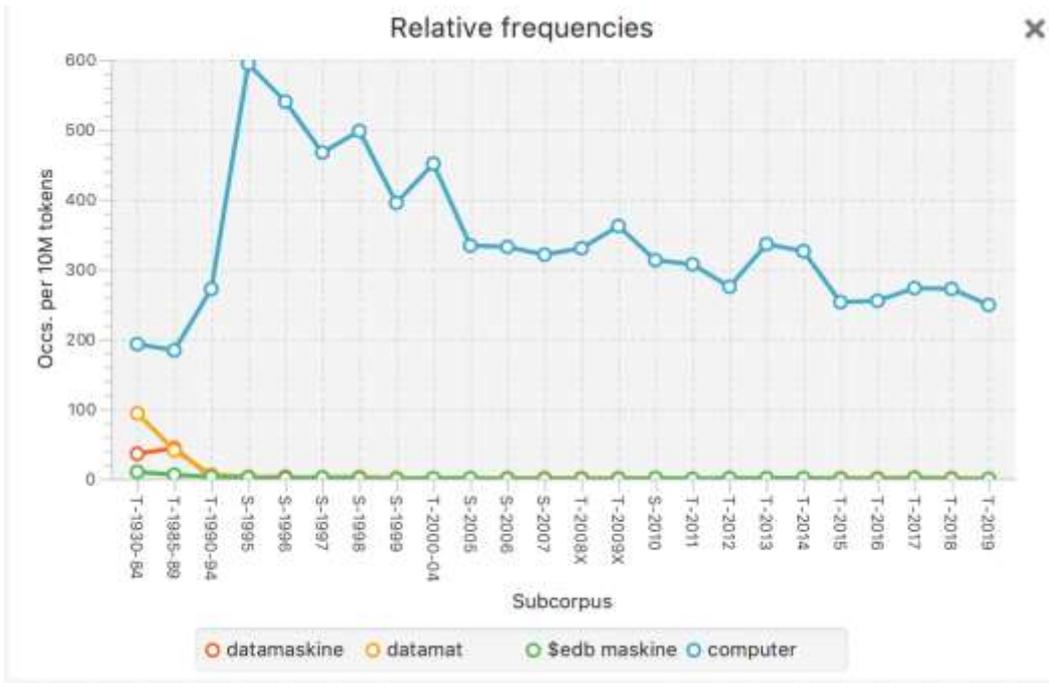


DIACHRONIC FREQUENCIES OF *GADEMAD* AND *STREETFOOD*

In the early days of computing, competing words for ‘computer’ existed, and Danish coinages such as *edb-maskine* (*edb* is an acronym for *elektronisk databehandling* ‘electronic data processing’), *datamat* (*-mat* as in the last part of *automat* which may be used as an independent suffix, cf. *laundromat*, *bankomat* and others) and *datamaskine* competed with the English direct borrowing *computer* for some time. However, since the mid-1990s, *computer* has dominated, as can be seen in Figure 2, and the Danish coinages are more or less obsolete today. This contrasts with Swedish *dator* and Norwegian *datamaskin*, both of which are language-internal word formations that are used much more widely in those communities than the English *computer*.

However, this particular development cannot be used to predict the development of other pairs. From the same domain, the words *homepage* and the Danish loan translation *hjemmeside*

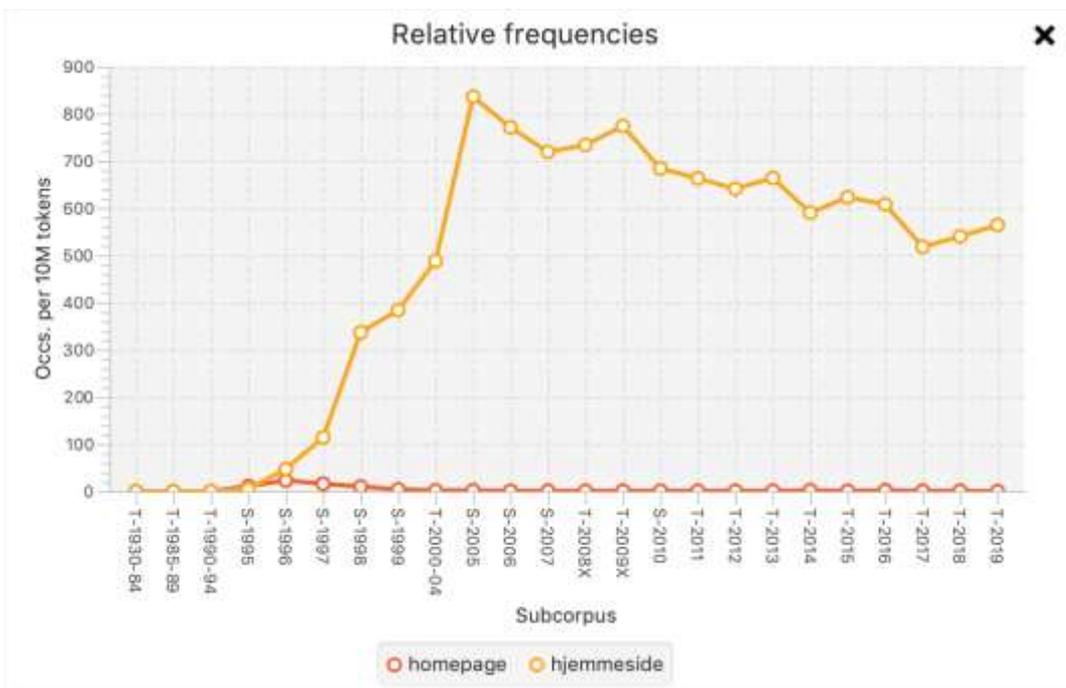
Figure 2



COMPETING WORDS FOR *COMPUTER* OVER TIME

were likewise in competition with one another in the mid-1990s. *Homepage* was in the lead until 1995–96, but from 1997 onwards, *hjemmeside* took over, and today the Danish word is totally dominant (Figure 3).

Figure 3



DEVELOPMENT OVER TIME FOR THE PAIR *HOMEPAGE* AND *HJEMMESIDE*

From the charts it is possible to trace the development of one or more words and use it for description in the dictionary. But it is difficult to generalize about the development of individual competing words. Sometimes both words co-exist for a long period of time. Examples include grammatical terms which have long existed in two sets: an international set based on Latin: *subjekt*, *objekt*, *infinitiv* etc. and a parallel Danish set with the same content: *grundled*, *genstandsled*, *navnemåde* etc. In other instances, one of the words becomes dominant and may eventually oust the alternative, but it is not possible to predict whether the internal or the foreign word will win.

The lesson to learn from this is that editors should be aware if more than one word is used for the same phenomenon. If a language-internal word co-exists with an English borrowing, it may be wise to include both, unless the frequency distribution is markedly skewed, so as to offer users a choice between them.

Pseudo-Anglicisms. One group of words that more clearly than others bear witness to the high prestige of the English language are the so-called pseudo-Anglicisms (cf. Gottlieb 2015): words that are clearly recognizable as English words, but have no parallel in the English language or are so infrequent in English that borrowing is an unlikely source. They are few in number, but still interesting from a language sociology point of view. In the material we find *flexicurity*, not only a pseudo-Anglicism but also a portmanteau, a contraction of *flexible* and *security*, and used to characterize the Danish labour market model which seeks to combine employers' need for a flexible working force with employees' need for job security. Other examples of neologisms in DDO are *badwill* (the opposite of *goodwill*), *citybike* (a type of bicycle for everyday purposes), *mooncar* (a kind of go-kart used by children in playgrounds) and *motherfuckerskæg* ('Vandyke or circle beard', blend with *skæg* 'beard').

Finding candidate words. The main reason why the numerical findings in this study should be treated with caution is that the headwords included were not randomly selected. Nor were they carefully selected according to predefined criteria of corpus frequency, domain distribution and other parameters. As demonstrated elsewhere (e.g. Trap-Jensen, Lorentzen and Sørensen 2014),

corpus frequency is less important for words beyond the first 100,000 headwords in the dictionary, and therefore the current selection of words may be biased in some way, for example as the result of the editors' conscious or subconscious preferences in the editorial process: they may be especially qualified to edit a particular group of words, or have given priority to words from certain semantic domains etc.

The editors are presented with various lists of candidate words from which they may select entries to edit (Figure 4). In the most recent period the staff focused in particular on user involvement and the editors deliberately selected more words from the list of user proposals than they would perhaps otherwise have done. It is likely that words proposed by users are in some way more conspicuous or remarkable. If this is true, one would expect Anglicisms and other words that stand out in some way or another to be over-represented. Conversely the inconspicuous, plain words that are just as frequently coined are more likely to go unnoticed by the average user, and for this

Figure 4

Word	Who	Weight
parathion	i-D...	477
anvender	i-D...	430
kompetencecenter	i-D...	417
lønnelpakke	i-D...	348
tuneserlov	i-D...	348
safety	i-D...	347
anlægsområde	i-D...	315
half	i-D...	303
evaluation	i-D...	286
walkie	i-D...	277
extension	i-D...	227
menneskebarn	i-D...	218
ædelæggende	i-D...	209
benævnt	i-D...	181

EXAMPLE OF WORDLISTS RANKED BY FREQUENCY FOR SELECTION BY EDITORS (BOX LEFT)

reason alone one should not attach too much significance to the proportion of different word-formation types in this study. An identical count over the next three-year period may show an altogether different distribution that can more likely be assigned to editorial practice rather than a real difference in the underlying language.

Conclusions

As mentioned above, the distribution in DDO between new and old words, between language-internal and foreign neologisms and between the different word-formation types should not be taken as facts about the Danish language. It says more about the dictionary and the lemma selection principles during the period of investigation. Having said that, the material does tell us a few things about neologisms and their role in the dictionary and in the language, and the status they have in the language community. Neologisms are important to keep a language dynamic and adaptive to an ever-changing world. At present, English is the dominant lingua franca of the world, and it comes as no surprise that this is reflected in the vocabulary when words for new things and phenomena come into the language, or that English words are used in slang and informal language even when language-internal words and expressions are available. From a language-systemic point of view, it does not matter if a word is an internal creation or an imported Anglicism, but it does matter to the people who speak the language.

DDO policy is to offer both entries if there is sufficient corpus evidence to warrant inclusion of competing words for new phenomena. A slightly different case is where a language-internal word does exist, but an Anglicism is introduced, often through teenage slang as a sign of the high status of English (*chick, hoodie, view* – not to mention perhaps the greatest English export success of all times: *okay*). Quite a few of these words are included in the dictionary, but the editors should be careful to make sure that corpus evidence is solid and not instances of code-switching in an increasingly bilingual society.

The analysis has also shown that new ways of forming words may be in the pipeline. Several instances of portmanteau words in the material could be an indication that this word-formation

process is becoming more common under the influence of English. The prestige of English is also clearly confirmed by pseudo-Anglicisms: English words created locally in Danish with no parallel in the implied source language. The empirical basis of the present study is too uncertain to determine if this is coincidence or a real trend, but it warrants further investigation.

References

- Bußmann, Hadumod. 1990. *Lexikon der Sprachwissenschaft*. Stuttgart: Alfred Kröner Verlag.
- Den Danske Ordbog* (The Danish Dictionary). <https://ordnet.dk/ddo>. Copenhagen: Society for Danish Language and Literature (accessed April 1, 2019).
- dialekt.dk* (website about Danish dialects and sociolinguistics). <https://dialekt.ku.dk>. Copenhagen: Copenhagen University (accessed April 1, 2019).
- Gottlieb, Henrik. 2015. Danish pseudo-Anglicisms: A corpus-based analysis. In *Pseudo-English. Studies on False Anglicisms in Europe*, edited by Cristiano Furiassi and Henrik Gottlieb, 59-98. The series *Language Contact and Bilingualism*, Vol. 9, edited by Yaron Matras. Berlin, Boston, Munich: Walter de Gruyter, Inc.
- Nye Ord i Dansk* ('New Words in Danish'). <https://dsn.dk/noid>. Bogense: Danish Language Council (accessed April 1, 2019).
- Oxford Learners' Dictionary*. <https://www.oxfordlearnersdictionaries.com>. Oxford: Oxford University Press (accessed April 1, 2019).
- Trap-Jensen, Lars, Henrik Lorentzen, and Nicolai Hartvig Sørensen. 2014. An odd couple – Corpus frequency and look-up frequency: what relationship? *Slovenščina 2.0*, Volume 2 (2014), Issue 2: 94-113, edited by Iztok Kosem and Michael Rundell. Ljubljana: Trojina, Institute for Applied Slovene.
- Wiktionary*. <https://en.wiktionary.org> (accessed April 1, 2019).