

Descriptive and Normative Aspects of Lexicographic Decision-Making: The Borderline Cases

Lars Trap-Jensen

The Danish Dictionary
Danish Society for Language and Literature
Christians Brygge 1
DK-1219 Copenhagen K
Denmark
ltj@dsl.dk

Abstract

The paper deals with that all-too-familiar situation where the lexicographer must make a decision whether to include a particular item in the dictionary or not, and if so then in what form. Some borderline cases are investigated where considerations of language use, system and norm are each important parameters in the attempt to reach a common understanding that can be used by the dictionary staff as a manageable guideline for making uniform decisions. Experience indicates that there is no obvious, unequivocal path that should be followed — indeed the outcome for one type of information may well be different from that of another as the subtle balancing of the aspects involved may vary and result in different solutions.

Introduction

Nowadays, any dictionary claims to be descriptive. Prescription of the norm for the good and correct language strikes an unpleasant note of elitism belonging to the past where the language of some privileged group could be elevated as an ideal to be followed by the more ignorant masses. In this paper, I shall look at some of the borderline cases involved in the dictionary-making process, requiring constant choices being made by the lexicographer in trying to decide whether to include a particular item in the dictionary. Any lexicographer has to deal with such decisions on a daily basis as I am sure many of the participants realize from personal experience. I am going to examine some areas where decisions of such a nature are particularly obtrusive, and elaborate the considerations that the staff has been through at The Danish Dictionary, leading to the results that can be seen on print. I shall do so from the perspective of the descriptive/normative dichotomy that has, in our experience, always been significantly involved whenever decisions on peripheral phenomena had to be made. The areas that I have chosen to look at are: orthography, meaning, inflection and lemma selection.

1 Normative and descriptive

My primary concern is practical, but I have to say a few words about the notions 'descriptive' and 'normative' in order to avoid confusion. For most practical purposes the two are regarded as more or less mutually exclusive terms: the dictionary articles either reflect language usage as observed from a text corpus or another collection of natural text production, or they reflect an externally given norm considered to be correct by some standard. However, I have found it useful to follow Nikula [1992] on one important

observation. She observes that descriptive and normative (or prescriptive which is here used in the same sense) should not be taken as proper antonyms as they refer to phenomena on two different levels. Description characterizes the relation between the lexicographer and the linguistic material at hand, be it a national language, a particular variety, or the language of a single speaker. Descriptive statements are by nature neutral or objective, whether the term is defined as inductive (e.g. in Bergenholtz-Schaeder [1978]), or verifiable or intersubjectively controllable (e.g. in Wiegand [1987]). From this follows that a considerable amount of information given in a dictionary does not lend itself to description proper as it is in essence evaluative rather than verifiable. For instance, the number of meanings and submeanings of a given lemma, labels such as 'informal', 'vulgar', 'derogatory', 'dated', or 'literary', or information on sense relations (synonyms and antonyms) cannot be judged true or false as they are inherently dependent on the lexicographer's analysis and his or her evaluation based on personal intuition. That alone indicates that there is no such thing as a purely descriptive dictionary. Normative on the other hand, still according to Nikula [1992], characterizes the relation between the lexicographer and the dictionary user. Prescription presupposes a norm – whether an officially authorized norm or a de facto norm that serves as a standard by which individual instances can be judged as deviations or violations. In the dictionary, norms can be explicit, e.g. when rendering the officially codified orthographic forms, or they can be covert, as when some linguistic items are included while others are excluded from the dictionary because they are judged recommendable or the opposite, irrespective of their frequency of occurrence.

With these points in mind, let us now turn to look at some of the problems we had to face during the making of our dictionary. The Danish Dictionary was commissioned to prepare a dictionary which should be "descriptive in the overall aim. It should, however, be normative with respect to information on orthography, and guiding with respect to information on inflection, word formation, collocations, and constructions" (quote from the official mandate). The areas that I have chosen to focus on are illustrative as they have a good deal in common when it comes to the empirical basis on which decisions are made, and on the other hand display different solutions reflecting the delicate balance between the aspects involved.

2 Orthography

The most clearly normative aspect of the dictionary is orthography. It is explicitly formulated in the mandate, and it is the only part of the language that is regulated by formal authority. In Denmark, the Danish Language Board has been given authority to establish the official spelling of Danish words, and these forms are of course followed throughout the dictionary. However, descriptive considerations forced us to think about frequent deviations from the norm, too, and eventually led to the inclusion of, as reference articles, a number of variant, norm violating forms, typographically clearly marked as unofficial. This is not an uncommon practice in dictionaries, the obvious pedagogical reason being to avoid the user looking up in vain because he/she is looking at the wrong alphabetical place. Somewhat more controversial is undoubtedly the decision to bring information on variant spelling in the main article as well. We have done so, not in the case of simple misspellings, but in those cases where the deviant form is motivated by a competing norm and therefore is perceived as less stigmatizing. This holds for instance with respect to some of the vocabulary of certain

professions, e.g. chemistry and medicine, where an international norm based on Latin is established among specialists in the field, but often in conflict with the official norm which may have been adjusted to the Danish spelling tradition. It also applies to a number of loanwords where the norm of the source language competes with an official, adjusted norm that has not yet become fully established in the language use. Our decision to include such forms runs counter to the recommendation of some lexicographers (e.g. Svensén [1987]) that deprecate such a practice on the ground that inclusion in itself renders a normative status to the unofficial form which is unfortunate and should at best be avoided. Svensén's recommendation gives priority to the normative aspect, while our decision allows more for the nature of the linguistic material we are describing, without, however, being entirely descriptive as we are, as it were, turning the unofficial form into a gradable: some forms are more unofficial than others, you could say.

3 Inflection

With respect to inflection, the norm is not quite so clear. In the official orthographic dictionary inflectional forms are given, but their status is of a somewhat different nature. Formally, the Danish Language Board does not have the same authority with respect to inflection, but in practice it hardly makes much difference. The fact that inflectional forms are included in the official dictionary makes the average user interpret the status of inflection on a par with spelling, it is probably fair to say. Correspondingly, we have chosen to include frequent, but deviating inflectional forms, although again clearly marked as unofficial. This is not different from the situation regarding orthography just mentioned. But in addition to this, one has to clarify what it means when an inflectional form is present: does it mean that it is frequent? Does it mean that it is attested? Or does it mean that it is possible, thinkable, or merely part of a general paradigm that traditionally applies to the category in question? When you analyze corpus data, you constantly find that inflectional forms given in the dictionary are in fact not attested in actual text productions, or the opposite situation of inflectional forms occurring that are not part of the official paradigm. The problem is primarily one of interpretation: is the absence of an inflectional form an indication that it does not exist, or is it an indication that the corpus is simply not large enough? And is one single instance of, say, a plural form of a noun out of a total of perhaps hundreds or thousands of occurrences sufficient to say that the noun is countable? Our conclusion is that it is not. Word inflection is an excellent example of an area where corpus evidence can be used to check and adjust the predictions of linguistic competence theory as discussed by Hanks [2000]. We have found it useful not always to bring the full paradigm automatically, but to check with corpora if the different forms are actually used. If forms were absent or of extremely low frequency, we have checked with our personal intuition, and in many cases chosen to omit the form in the dictionary. The result is consequently more in accordance with language use, showing the normal, widespread use, rather than what forms are possible as seen from the language system point of view. In my view it is a sound decision, although some people may be disappointed (or maliciously amused) to find logical gaps or inconsistencies, for instance when a compound noun is given an inflection which differs from that of the simplex word.

4 Meaning

We all know that the uniform picture of a language given in language courses and sometimes also in dictionaries is a postulate that may have little to do with linguistic reality. Language is intrinsically non-constant, it is dynamic and it changes. Sociolinguistics has taught us that variation and variability are not just surface phenomena that can be discarded as accidental performance errors muddling the neat, logical system of competence. No, variation is part and parcel of the dynamic nature of language. Indeed, diachronic change manifests itself synchronically as variation within the speech community. And this is precisely what we find when we analyze performance data. Some of the areas most susceptible to variation are vocabulary and meaning (leaving aside pronunciation in this connection). As far as meaning is concerned, variation reflects the fact that the contents of a particular word are not shared by all members of the language community. Common examples are technical meanings used primarily by specialists, older meanings gradually fading away, or new meanings that are evolving in, say, the language of younger people, but which are not yet shared by the larger community of speakers. The usual way for lexicographers to deal with such instances is to include them (descriptively) in the dictionary, but supply them (normatively) with a label that indicates the nature of the restriction on their usage. This works well and is by and large unproblematic as most people are prepared to acknowledge the existence of a meaning of which they had so far been unaware. There are, however, cases which are more controversial and problematic, and typically involving meanings which are mutually incompatible, or which are judged 'illogical', 'incorrect' or downright 'wrong' by some or most parts of the language community. Unlike the first examples these meanings do very often not go unnoticed, but are made object of public debate and by many people seen as examples of linguistic decay etc. There are different solutions to the problem that present themselves. One solution is simply to neglect the existence of such usage and exclude the meanings from the dictionary, but for us it was not an attractive solution as it would violate the descriptive mandate. Alternatively, one could simply include the meaning together with the other meanings of the word without further fuss, arguing that any meaning of sufficient distribution should be given equal treatment in the dictionary. However, we felt that a sign of warning would be appropriate in order to prevent users from unintentionally being classified as 'poor language users'. So what we have done is to include the meanings if they are worthy of description according to the usual criteria, and supplying them with a label 'this usage is regarded by some/many as incorrect' signalling the controversial nature of the meaning. This solution is related to the usage notes that are becoming increasingly widespread in modern dictionaries (e.g. in the NODE), in my view to the benefit of the user seeking guidance on controversial issues. The label can of course also be used in connection with other information types that are being used in a similarly disputable way, e.g. valency patterns, constructions with one preposition rather than another or one lexical or inflectional form being substituted for another.

Although the deliberately neutral wording of the label may make the lexicographer feel descriptively objective, it is important to realize that the label is still a normative statement. The fact that the lexicographer disclaims personal responsibility for the incorrectness of the meaning does not make it less normative, and neither does it make it normatively neutral. Its controversial nature is still seen from the point of view of the educated, unerring and well-

established language users. You cannot guide and make recommendations and at the same time remain normatively neutral.

5 Lemma selection

Within vocabulary there is one area which is particularly important as seen from the normative/descriptive point of view. Denmark - like many other countries - experiences in these years a massive pressure from the dominant Anglo-American culture, and a part of this pressure is linguistic. There is a constant flux of new terms and loan words from English entering into Danish, Danish speakers are becoming increasingly Danish-English bilingual, and in some environments, especially juvenile subcultures, it is regarded as cool to use English words or even entire sentences embedded in an otherwise Danish discourse. From this follows that a substantial number of English words are to be found in Danish texts, and when your lemma selection procedure is partly based on texts from a balanced language corpus you are obliged to consider a good deal as candidates for inclusion in the dictionary. The issue is not only sensitive in terms of language policy, it has also proved extremely difficult to reach internal agreement within the staff on the right analysis and interpretation of their status as a precondition of eligibility in the first place. It is not so much a matter of linguistic tolerance vs. purism. Danish has a tradition of being fairly open to linguistic borrowing, at least in comparison with the other Nordic languages, and it is not an issue here whether words like swimmingpool, weekend or computer should be included in the dictionary. They are well-established as Danish words and no-one would seriously dispute their legitimacy as lemmas. No, what we are talking about here can by and large be organized in two major groups. On the one hand, there is a large number of terms and specialist words that are being borrowed, very often together with the phenomenon itself e.g. belonging to science or professional areas, sports, music and other cultural areas, and on the other there are words and expressions from the common English language being used first and foremost by young people, often as a social identity marker of the subgroup or generation and associated with the high status which American English enjoys in these environments. Within the first group one finds many words that are in the centre of heated language debates in various countries, and as such it is a far more complex topic than I can go into in this context. Suffice it to say that we have treated words from this group little different from the rest of the vocabulary, i.e. they have been included if we have found sufficient empirical support in our corpus and the other sources that we used. In cases of doubt, it is worth considering if the candidate word shows any indication of adopting Danish morphology and pronunciation, and also if a competing Danish coinage exists. In the latter case we either include two articles, a main article and a reference article, or one article with the competing form given as a synonym, and generally with the more frequent word as the main article. Only in the case of both overall low frequency and near-equal numerical distribution we have, normatively, given the Danish word priority over the English. With respect to the second major group it has proven notoriously difficult to reach agreement on the analysis and interpretation of data themselves. Some hardcore descriptivists maintain that any word which is attested in Danish discourse should be considered as a lemma candidate, and the only valid criteria for inclusion is frequency and distribution. Others argue that what we observe are instances of code switching, that is words or phrases of one language being inserted during the production of a text in another language. Personally I have little doubt that clear examples of code switching have no place in a dictionary of the other language,

but I admit that it is extremely difficult to distinguish objectively between borrowing and code switching in natural texts. Our experience is that the more used you are to hearing and perhaps using English words in your own conversation, the more inclined you are to accept them as belonging to the vocabulary of Danish.

Concluding remarks

Improved access to large text corpora has no doubt provided lexicographers with a more solid foundation for making statements about language usage and has prompted a welcome shift in focus from the categorical possibilities of language competence theory towards a more probabilistic view derived from observations of actual language performance (cf. [Hanks 2000]). Some of the borderline cases involved in the inflectional paradigm suggest where this shift may take lexicographic practice. Likewise, the other areas investigated illustrate how linguistic diversity can be catered for in the dictionary as a result of descriptive considerations. However, from this does not follow that a corpus based dictionary automatically becomes a descriptive dictionary. On the contrary, it is a fallacy to think that a dictionary can be entirely descriptive. On the one hand, dictionaries contain information that does not lend itself to description at all, and on the other hand, lexicographers still have to take other aspects than usage into consideration. Dictionary users expect guidance with respect to correctness, and this in turn requires the existence of a norm. Lexicographers need to recognize their responsibility in the process of establishing language norms: dictionaries not only reflect a particular linguistic state, they also play an important and active role when it comes to deciding what constitutes the norms of a standard language.

References

- [Bergenholtz-Schaeder 1978] Bergenholtz, H., Schaeder, B., 1978. Ausblicke auf eine deskriptive Lexicographie, in: *Interdisziplinäres deutsches Wörterbuch in der Diskussion*, pp. 116-172, Sprache der Gegenwart Band XLV, Düsseldorf.
- [Hanks 2000] Hanks, P., 2000. Contributions of Lexicography and Corpus Linguistics to a Theory of Language Performance, in: *Proceedings from The Ninth EURALEX International Congress vol. I*, pp. 3-13, Universität Stuttgart, Germany
- [Nikula 1992] Nikula, K., 1992. Deskriptive ordböcker – finns dom?, in: *Nordiske Studier i Leksikografi no. I*, pp. 43-52, Nordisk forening for leksikografi, Oslo
- [NODE] The New Oxford Dictionary of English, 1998, Clarendon Press, Oxford
- [Svensén 1987] Svensén, B., 1987. Handbok i lexicografi. Principer och metoder i ordboksarbetet, Esselte Studium, Stockholm
- [Wiegand 1986] Wiegand, H.-E., 1986. Von der Normativität deskriptiver Wörterbücher. Zugleich ein Versuch zur Unterscheidung von Normen und Regeln, in *Sprachnormen in der Diskussion. Ein Beitrag vorgelegt von Sprachfreunden*, pp. 72-101, Walter de Gruyter, Berlin, New York