Inflection and Word Identity

Maarten Janssen
IULA, Universitat Pompeu Fabra, Barcelona

This article argues that inflection should be seen as a (partial) criterion that defines homonymy: when two word meanings have different inflected forms, they have to belong to different lexical entries. If this were not the case, it could not be maintained that inflection is a property of lexical entries, but we have to rather say that each word sense has its own inflectional paradigm, even though in most cases all senses of a word inflect in the same way. Although there are apparent cases where it looks like inflection might be in fact dependent on word meaning, none of these cases really goes against the hypothesis that inflection is a property of lexical entries, and not of word senses.

1. Introduction

An obvious fact of language is that words inflect, at least in inflectional languages. However, the word 'word' is one of the least well-defined notions in linguistics. An important question in inflectional morphology is therefore: what exactly are the things that inflect? It is clear that what inflects is not a word-form: word-forms are the result of inflection, not the source of it. That leaves two obvious candidates: lexical entries or word-meanings. In this article I will explore the question which of these two options should be considered correct.

The question what inflection applies to is not a question that is often addressed, at least not in lexicon design. But any lexical model that features inflectional forms has to deal with this question one way or another. This article is written in the light of the Open Sources Lexical Information Network (OSLIN, Janssen 2005). OSLIN is a full-form lexical database model, which has three main levels: a level of word-forms, lexical entries, and word-senses. Each word-sense is linked to a lexical entry, and in this model, the issue at hand boils down to the question whether word-forms should be linked to lexical entries or to word-senses. However, the discussion in this article equally applies to the role of inflection in lexica in general, including their role in dictionaries.

There are several linguistic models in which inflection is linked directly to word-senses. Maybe the most explicit of those is the Meaning-Text Theory (Mel'cuk 1981). In MTT, inflection is applied at the level where words have already assumed a specific sense. And therefore, in MTT the *lexemes* (or sense-specific words) carry inflections. When inflection is treated separately for each word sense, then inflection can be represented correctly independently of the question whether inflection is or is not word-sense specific. However, treating inflection at that level does have the disadvantage that inflected forms have to be repeated for each sense of a word.

The number of theories in which inflection is word-sense-based is relatively small: the majority of linguistic theories link inflected forms to words or lexical entries, as do the majority of lexicon models. This leads to a much more efficient representation of inflected forms, where each form is shared by all senses of a word. However, there are several (apparent) counter examples in which this approach leads to incorrect predictions, as will be discussed below.

Dictionaries typically take a middle stance: the majority of inflectional information is represented at the level of the lexical entries, but a small number of inflectional indications can be indicated at the level of the word meaning. Whereas this is a pragmatic solution, it is not very attractive as a linguistic model: it basically forces you to treat inflection on both the

level of lexical entries and the level of word senses, since inflectional information can be found at either level. Since that it hardly a workable position, the solution adapted in dictionary should rather be seen as an economic shorthand: inflection always takes place at the level of word-senses, but whenever all word-senses happen to share their inflectional paradigm, the inflected forms are represented only once, at the level of the lexical entry. Yet that indication at the level of the dictionary entry should be taken to mean that that inflectional information applies to every word-sense under it.

This article defends the claim that lexical entries inflect, and not word-senses, in spite of the apparent counter examples. The proposal is implemented by making inflection part of the identity criterion of lexical entries: whenever two word senses inflect differently, they cannot belong to the same lexical entry. Although this seems a rather innocent claim, it does have some serious consequences. However, I argue that these consequences are acceptable in the light of the objective.

2. Identity Criteria

If we assume that inflection is a phenomenon that takes place at the level of lexical entries, then inflection cannot be dependent on word meaning. That is to say, if the inflectional paradigm belongs to the lexical entry, then it cannot be the case that the same word used in a different meaning inflects differently: the plural of *ball* is *balls*, and that has to be independently of which meaning the word *ball* is used in. We cannot even speak of the inflectional forms or the inflectional paradigm of a word sense in that case, since word senses do not have inflections: word senses can only be said to *invoke* inflection: a word sense invokes an inflectional paradigm when it belongs to a lexical entry that has said inflectional paradigm. Therefore, two word senses that belong to the same lexical entry necessarily invoke the same inflectional paradigm. Summing up, in order for inflection to be a property of lexical entries, hypothesis 1 below has to hold, because if it does not, inflection has to be attached to word senses.

> *Hypothesis 1* Two word-senses that invoke different inflectional paradigms can never belong to the same lexical entry.

To point out several obvious things which are not required by hypothesis 1. Firstly, it is of course possible for the same word to have different plurals in different meanings: the word *band* (tyre; band) in Dutch has a plural *banden* in its meaning of 'tyre', but *bands* in its meaning of a musical band. However, such distinction can only exist in cases of homonymy. In other words, if hypothesis 1 holds, then inflection distinguishes polysemy from homonymy: whenever a word has a different inflection in different meanings, it has to be homonymous.

Secondly, it is possible for a word to have only one inflectional paradigm and still be homonymous: the plural of the word *sole* is *soles* in all of its meanings, but that does not mean that there has to be only one lexical entry for it. In other words, hypothesis 1 does not entail that only different inflection leads to homonymy. Due to the regularity of the inflectional system in English, most homographs will in fact inflect in the same way.

Finally, it is not a consequence of hypothesis 1 that whenever a word shows variation in its inflection, it has to be homonymous. In the case of nouns: when a word has two different plurals, it does not necessarily have to be homonymous. Take the Portuguese word *vulcão*

(volcano; eruptive situation). It has two different plurals, either *vulcões* or *vulcãos*, yet the word *vulcão* is polysemous and not homonymous. However, for that to be possible, the variation has be a property of the (entire) lexical entry and be independent of word meaning. In other words: both plurals have to be available in each meaning of the word *vulcão,* as is indeed the case. Independently of whether you are talking about a number of volcanoes or a number of 'eruptive situations', both *vulcões* and *vulcãos* can be used as a correct plural. And in the same sense, you find a meaning-independent inflectional variantion in the past tense of the Dutch verb *waaien* (to blow – either *waaide* or *woei*), as well as in the plural of the English words *cactus* (either *cacti* or *cactuses*), *corpus/corpora, bureaus/bureaux, compendiums/compendia* , *drachmas/drachmae*, *volcanoes/volcanos*, etc.

What is implied by hypothesis 1, however, is that if a word has a different inflectional paradigm in different word-senses, then that word has to be homonymous. That is to say, if there is variation in the inflectional paradigm of a word that is correlated to the meaning in which the word is used, then there have to be two separate lexical entries for that word. In the example of the Dutch word *band* above, what is the correct plural depends on the meaning the word is used in: you cannot use the plural form *banden* when referring to several groups of musicians, and neither can you use the plural form *bands* when talking about tyres. Therefore, there have to be two homonymous nouns *band* in Dutch.

If there are cases in which hypothesis 1 does not hold, we will be forced to conclude that inflectional paradigms belong to word-senses and not to lexical entries. In that case, it would have to be the case that each word-sense carries its own inflectional paradigm, even if the paradigm for each sense of a word is almost always the same. It would also mean that in principle, the preface of the dictionary should state that even though the inflectional information is given directly under the headword, that that is just an efficient way of presenting the inflectional paradigm of each word sense, but that strictly speaking, the inflectional information should have been presented for each word sense.

There are several arguments around in favour of the claim that hypothesis 1 cannot, in fact, be correct. The remainder of this section will discuss the two main ones, and try to defend hypothesis 1 against them. The first argument has to do with defective paradigms; the second with words that have meaning-dependent plurals yet are taken not to be homonymous.

2.1 Partial Defectiveness

There is at least one large class of words that seem to have a meaning-dependent inflection, namely the words that can be used both as mass and as count nouns. Take for instance the two word-senses of the word *port*: in its reading of 'a type of fortified wine' it is a mass noun, which does not have a plural form. Whereas in its reading of a 'harbour' it is a count-noun, and has the plural form *ports*. Lyons (1977) quotes this difference in grammatical subclass as a reason why *port* has to be homonymous. In the Meaning-Text theory, this is a commonly quoted motivation for the fact that words that have both mass and count readings, including deverbal event nouns such as *influence*, have to have two separate lexemes.

If hypothesis 1 is taken strictly, that means that there have to be two lexical entries for *port* and all words alike. In the best case, that would mean that a large number of nouns would require more than one lexical entry. And if you take into account the regular possibility to shift between mass and count noun, as for instance in the well-known example '*There is cat all over the driveway*', it would imply a dual entry for basically every noun in the lexicon.

However, there is a crucial distinction between the difference in the plural forms for nouns like *band,* and the difference in the presence/absence of a plural in the case of words like *port.* The difference between the inflectional paradigms of the two senses of *port* is not that they contain different forms, but that there are forms missing in one of them. That makes it possible to say that they *do* in fact have the same inflectional paradigm, but that not all forms are usable in each meaning. In other words: if the mass noun reading of *port* would be usable in plural form, its plural would have to be *ports*, as in fact it is: just look at sentences like *'Rubies and tawnies are the least expensive ports made.'* It is only because of a specific semantic (or grammatical) restriction that as a mass noun, the plural form cannot (easily) be used.

In the present paper, this phenomenon is referred to as the *partial defectiveness* of a paradigm. I claim that the paradigm of the lexical entry *port* has both a plural and a singular form, and the plural and singular apply to all senses under the lexical entry. But there can be independent restrictions on whether or not individual forms can be used. This means that a small correction to hypothesis 1 is needed: different senses of the same lexical entry have to invoke the same inflectional paradigm, but they do not have to invoke it completely. In some of the word-senses, the inflectional paradigm can be defective.

Partial defectiveness does not only occur in the case of mass/count nouns but also, for instance, in the case of verbs that can be used both in impersonal and personal constructions. A nice example is the Portuguese verb *chover* (rain) which has a 3rd person plural *choveram*, which cannot be used in the standard meaning of 'rain', but only in metaphorical sentences like '*choveram as criticas*' (it was raining criticisms).

Partial defectiveness is more than just a pragmatic trick to rescue hypothesis 1: it is an independently motivated phenomenon. The partial defectiveness of *port* occurs not because there are special inflectional rules for one of its senses, but due to semantic blocking. It is the fact that in its mass noun reading, *port* semantically refers to something that cannot be referred to in the plural. And therefore, the plural form cannot (easily) be used in that specific meaning of the word. Therefore, partial defectiveness is only expected in cases of semantic blocking, and not in cases where the defectiveness is due to other reasons. For instance, there is a group of verbs in Portuguese that are defective for historic reasons, including the verb *ressequir* (to dry out). These verbs cannot be used in the first person singular present indicative (as well as some other forms), but this restriction has no semantic grounding. And as would be expected, in those cases the defectiveness is never partial: it always affects all the meanings of the verb.

2.2 Meaning dependent inflections

There is a small number of cases in English in which there is a single word, which is supposed not to be homonymous, yet according to the dictionary has a different plural depending on the meaning the word is used in. One of the clearest examples of this is the English word *appendix*, which is commonly taken to have a different plural for the medical and the general language meaning, as illustrated by the excerpt from the Longman Dictionary of Contemporary English (LDOCE) in figure 1.

ap•pend•ix *noun or* [countable]
1 *plural* appendixes a small organ near your BOWEL, which has little or no use:
Christine had to go into hospital to have her appendix out (=have it removed).
2 *plural* appendices a part at the end of a book containing additional information:
See Appendix 2.6

Figure 1. Definition of *appendix* from the LDOCE dictionary

Taking the inflectional criterion to word identity seriously, one would have to say that the lexical entry in figure 1 is faulty, and that it is so for one of the following reasons: (1) the plural forms of substantives are not part of the inflectional paradigm; (2) there is in fact no difference in the plural of the two meanings; (3) the two different readings are linked not to the meaning, but to the domain or register; or (4) the two meanings are not (no longer) part of the same word. Although the first and third of these options are not very appealing, both the second and the fourth option provide an appropriate answer, as will be argued below.

Option 1: It is possible to argue that the inflectional paradigm of a noun does not contain the plural, but that rather the plural form is a derived form of the singular. If that were true, then the two different plurals for *appendix* would not go against hypothesis 1 since they are not part of the lexical paradigm that is invoked. However, firstly, doing so would go strongly against a long-standing tradition, even though for instance Booij (1996) argues that the plural is 'only' a case of inherent inflection. But secondly, it is not only an unattractive solution to say that the plural of a noun is not inflectional, it does not even solve the problem: though this might be a way out for plural forms, there are similar examples from different word classes, such as the past tense of the verb *hang*, which is an irregular verb (*hung*) in its common meaning(s), but a regular verb (*hanged*) in the meaning of 'executing by hanging'. Solving all the examples by limiting the scope of the inflectional paradigm would basically mean getting rid of inflection altogether.

Option 2: Even if normatively there is a difference between the plural of the word *appendix* in its different readings, one could argue that that difference does not exist, or at least not exist as strongly, in the actual use of the word. An Internet search shows a large amount of cases in which the word *appendix* is used with the 'wrong' plural that does not correspond to the appropriate meaning (according to the dictionary). There are many cases in which the non-medical term is used with the regular plural: *This letter includes two appendixes.* There are also cases where the Latin plural is used for the medical reading of the word: *For what reasons are appendices removed?*

However, there are two problems with saying that there is in fact no difference in the plural of the two senses of *appendix*. In the first place, it ignores all the lexicographic evidence that was used to indicate this difference in the dictionary in the first place, not only those based on acceptability issues, but also those based on frequency counts. And in the second place, it is not a very satisfactory response to the problem at hand: if the two meanings of *appendix* have the same plural, that would merely mean that *appendix* is in fact a bad example of a word with a meaning-dependent inflectional paradigm. And unless we claim that there is also no difference in the forms of all other reported cases of meaning-dependent inflection (see for instance table 1 below), the problem would remain in the case of other examples.

Option 3: There are cases where inflection is region-specific. For instance, in Catalan, the plural of *home* (man) is *hòmens* in the Western variant, but *homes* in the Eastern variant. And the fact that in different areas the plural is formed differently does surely not imply that *home* is homonymous. Along the same lines, one could argue that the apparently meaning-

dependent use of the plural form *appendixes* has nothing to do with the meaning in which it is used, but rather the correct plural specific to the language of the domain of discourse. However, if that were true, you would expect both plurals to behave similarly in the same domain: in a medical text, the plural *appendixes* should be used for both meanings of the word, whereas in non-medical text, the plural *appendices* should be used even when it refers to the internal organ. And that expectation is not in line with the actual data: the correlation is with meaning, and not with the domain of discourse.

Option 4: When all other options fail, there is only one option left: the two words *appendix* might be etymologically related, but synchronically, they do no longer behave as the same word. It is not uncommon for a single word to gradually grow into distinct words in different meanings. As an example, one could consider the Portuguese word *estória* (story), which is in principle the same word as *história* (history), but written slightly differently. Tradionally, these two words were different senses of a single homonymous word *história*, where one of the meanings grew to have an adapted spelling, and therefore the two senses grew into two different words.

In cases of orthographic differences like in the case of *estória/história*, it is uncontroversially accepted that the two word meanings no longer belong to the same lexical entry. The claim I defend here is that in the case of *appendix*, the difference in inflectional paradigm is equally indicative of the fact that the two word senses are living their separate lives, with the only difference being that contrary to the case of *história*, the two words *appendix* still share the same citation form. And indeed, there are dictionaries such as the *Cambridge Advanced Learner's Dictionary* (CALD) that list the two meanings of *appendix* as different lexical entries. Table 1 presents a list of words in LDOCE which are listed in the dictionary with a meaning dependent plural, which should be treated on the same footing as *appendix*.

Noun	Plural 1	Plural 2
appendix	appendixes (bodily organ)	appendices (part of a book)
cherub	cherubim (angel guarding God)	cherubs (all other meanings)
brother	brothers, brethren (member of religious group)	brothers (all other meanings)
antenna	antennae (part of an insect)	antennas (metal rod)
craft	craft (boat)	crafts (job)
grouse	grouse (animal)	grouses (complaint)
louse	lice (animal)	louses (nasty person)
mouse	mouses (computer device)	mice (all other meanings)

Table 1. Words in LDOCE with a meaning-dependent plural

For all the words in table 1 that are included in the relevant meanings in CALD, they have indeed received separate entries. It should be noted, however, that CALD does not always make this split on the basis of inflection in these cases: the computer-related word *mouse* is presented in an entry separate from the word for the animal, despite the fact that CALD indicates the plural *mice* for that meaning.

The list in table 1 is not the complete list of words for which a meaning-dependent plural is given in LDOCE; there are other words that would probably deserve a different treatment. These are: *buck, foot, grand, index, salmon, penny, person, pike, score, shark, skate, staff, stone, swine, ton,* and *trout*. A complete discussion of all these words is beyond the scope of this paper, but for all these words, it can be argued that they are not really cases of non-homonymous, meaning-dependent inflection.

3. Conclusion

As has been argued in this article, inflection should be seen as a (partial) criterion that defines homonymy: when two word meanings have different inflected forms, they have to belong to different lexical entries. This claim is motivated by the fact that if it were not true, it could not be maintained that inflection is a property of lexical entries, but have to rather say that each word sense has its own inflectional paradigm.

The claim that two meanings can never belong to the same lexical entry if they inflect differently has to be softened by the introduction of partial defectiveness: words can be partially defective, that is to say, words can lack certain inflected forms in some of their meanings. Partial defectiveness is independently constraint by semantic blocking, where semantic or grammatical features of the word prevent certain forms from being used.

From the perspective of traditional lexicography, the fact that words like *appendix* should be seen as homonymous does not pose a major problem: not only are there dictionaries such as the *Cambridge Advanced Learner's Dictionary* in which words like *appendix* and *hang* already have different entries for the different meanings, but modern dictionaries often group meanings under a single heading that clearly do not really count as the same word. For instance, it is not uncommon to list verbal and nominal readings of the same word in a single entry, despite the fact that they do not even belong to the same word class. But it should be made clear that the plural forms *appendices* and *appendixes* do not in fact belong to the same word.

For modern online dictionaries, the issue of where inflection belongs becomes more relevant than it used to be in traditional paper dictionaries: many online dictionaries provide verb conjugations, and for instance the *Grand Robert* furthermore provides the inflection for all lexical entries. The inflectional paradigm typically pops up in a separate window – but these windows always relate to the entry, and never to a word sense. Therefore, keeping inflection as a criterion for homonymy becomes much more crucial in modern lexicography.

References

Booij, G. (1996). 'Inherent versus contextual infection and the split morphology hypothesis'. In Booij, G. & van Marle, J. (eds.). *Yearbook of Morphology 1995*, Dordrecht: Kluwer.

Janssen, M. (2005). 'Open Source Lexical Information Network'. *Third International Workshop on Generative Approaches to the Lexicon, Geneva, Switzerland.*

Lyons, J. (1977). *Semantics.* Cambridge: Cambridge University Press.

Mel'cuk, I. A. (1981). 'Meaning-Text Models: A recent trend in Soviet linguistics'. In *Annual Review of Anthropology*, vol. 10, 27 - 62.

Dictionaries

Cambridge Learner's Dictionary. Third edition. Cambridge: Cambridge University Press, 2007.

Grand Robert, version electronique [on line]. [Paris: le Robert]. http://gr.bvdep.com.

Longman Dictionary of Contemporary English. Second edition. Essex: Longman, 1987.