Between Inflection and Derivation
Paradigmatic Lexical Functions in Morphological Databases

Maarten Janssen
ILTEC, Lisboa, Portugal
maarten@janssenweb.net

Abstract
This article shows the use of lexical functions in the MorDebe lexical database system. Lexical functions are used in MorDebe to bridge the gap between derivation and inflection – to model relations between word-forms that are derivational from one perspective but inflectional from another. However, lexical functions in MorDebe range over lemmas rather than word-meanings and are hence significantly different from the normal lexical function.

Абстракт
В этой статье обсуждается использование лексических функций в лексической базе данных MorDebe. Лексические функции используются в MorDebe для того, чтобы сократить разрыв между словообразованием и словоизменением – для моделирования отношений между словоформами, которые можно рассматривать с одной точки зрения как деривационные, а с другой - как словоизменительные. Однако, лексические функции в MorDebe определяются скорее для лексем, т.е. словарных входов, а не для смысловых единиц и, следовательно, отличаются от лексических функций в обычном понимании.

1. Introduction

Despite its central role in linguistics, the distinction between inflection and derivation is far from clear-cut. Modern day morphological theories often posit in-between cases - behaving partly like inflections, but partly like derivations. However, the introduction of marginal cases is problematic from the perspective of morphological database: an actual database which lists inflections and/or derivations, either has to treat them on a par, or model any individual form either as derivational or inflectional. This article will show how this problem manifests itself in the MorDebe morphological database, and how lexical functions are used to (partially) overcome the problem.

This article consists of three parts: the first part will show the set-up of the MorDebe database and its treatment of derivation and inflections. And the second part will illustrate the problem of distinguishing inflection from derivation from the perspective of the database, and how lexical functions are used to remedy this. But the application of lexical functions in this way implies a fundamentally different use of lexical functions: lexical functions as ranging over lemmas rather than word senses. The third part of the article will show how this affects lexical functions, and discusses the problems of this use of lexical functions.

2. MorDebe, OSLIN and SIM*u*LLDA

MorDebe is a large volume morphological database, set up language independently, but currently only used for (European) Portuguese. The database currently contains some 125.000 Portuguese

lemmas, and around 1,5 million word-forms. MorDebe is integrated with a semi-automatic neologism detection system called NeoTrack, which keeps the database up-to-date and ever expanding. MorDebe models only formal properties of word - and currently contains no semantics. But the MorDebe database is designed as the formal part of the SIM*u*LLDA system (Janssen, 2002): an interlingual lexical database system focusing primarily on semantic content.

SIM*u*LLDA has a layered set-up, distinguishing different lexical levels. The basic lexical unit is roughly the headwords in dictionaries – referred to as the lemma[1]. All lemmas are related to a set of inflectional forms on the one hand, and a set of word-senses or meaning on the other. The lemma itself is strictly an abstract, formal unit, adorned with a citation form mostly for practical purposes.

The lemma is defined by its inflectional paradigm – if two lemma have a different inflectional paradigm, they have to be considered different lemmas, as in the case of the Dutch noun *band* which can have *banden* (tires) or *bands* (bands) as its plural – or the Portuguese verb *redar* which can be inflected as *redou* (I give again) or *redo* (I catch with a net). However, within an inflectional paradigm there can be alternative forms: the first person singular present tense of *ouvir* (to hear) can be either *ouço* or *oiço*. The word-forms in the inflectional paradigm have an orthographic representation, an inflectional code, and possibly a dialectic restriction.

Every lemma can express one or more word-senses. The word-senses are stored in a separate database, explicitly linked to a lemma by means of unique identifiers. Apart from meanings, it is possible to relate other kinds of lexical information to the MorDebe database in the same fashion: pronunciation data, morphological, grammatical descriptions, etc. By linking all these different types of information to the same set of lemmas, the different databases form a network of lexical information, in which every lexical resource could be maintained by a different institute. At the moment, the idea using MorDebe for distributed lexical databases is being developed into a proposal for an open source lexical information network (OSLIN).

One of the databases currently under development for Portuguese is a database of derivations. Derivations are modelled as relations between lemmas. Like inflectional forms, derivations are stored explicitly rather than being rule-based generated. The fact that inflections are stored as part of the lemma, whereas derivations are modelled as relations between different lemmas implies that there is a clear distinction between the treatment of derivations and inflections in MorDebe. In the case a a number of morphological relations, this leads to a problem, since certain forms should be considered inflectional from one perspective, but derivational from another.

3. Between Inflection and Derivation

Inflectional morphology and derivational morphology are two traditionally distinct fields, attributed many types of differences - such as the claim that derivation is pre-grammatical, whereas inflection is part of the grammar itself (Split Morphology, Perlmutter 1998). Various distinctive traits have been formulated by which inflection can be distinguished from derivation. A good summation of the distinctive treats between inflection and derivation is given by Booij (*to appear*): (1) inflections are obligatory and fully productive, (2) inflection does not change the

[1] In SIM*u*LLDA, the basic entries are actually referred to as *lexemes*, whereas the *lemma* is the set of word-forms under the lexeme – but to avoid confusion with the sense-dependent notion of lexeme in MTT, the term *lemma* is used in this article.

lexical category, (3) derivation has no impact on grammar, (4) inflection is paradigmatic, (5) inflection has no semantic change.

But it has been argued by many authors, including Schultink (1962), Bybee (1985), and Booij (1995), that none of the proposed criteria define a solid division between derivation and inflection. Proposed solutions to this lack of clear distinctions include the introduction of an in-between category (Booij, 1995), or a class of in-between cases (Bauer, 2004), viewing the distinction as a gliding scale (Bybee, 1985), or drop the distinction altogether (Lieber, 1980).

This section does not discuss the differences between inflection and derivation in detail, but mainly illustrates the problems of a clear distinction from the perspective of that which is most relevant for lexical databases: inflection creates word-forms, whereas derivation creates new lemmas.

3.1. The derivational nature of inflection

There are clear examples of inflectional forms: gender and number on adjectives is inflectional in every sense of the word - the realisation of the adjective is completely determined by the noun it applies to (which is why Booij 1995 calls this *contextual inflections*), and the form of the adjective itself has no semantic, syntactic, or even pragmatic implication. But many other inflectional forms are less clear – those that Booij (1995) calls the *inherent inflections*: inflection that is "*syntactically relevant but not constraint by sentence structure*" (Bauer, 2004).

Inherent inflection has many derivation-like features. Inherent inflection forms often have a clear lexical meaning. For instance, female forms of nouns refer to female rather than male entities; if you play a piano piece *piano* it is not the same as when you play it *pianissimo*. And inflection is said to always relate to lemmas – not to meanings, but often inflectional forms only apply to specific meanings of a word: the female form *amarela* is only related to the word *amarelo* in its meaning of *pale person*, not in its more common meaning *yellow*.

Inflection should be paradigmatic, but diminutives, plurals and female forms can stack: the form *pequeninas* is the plural of the female form of the diminutive of *pequeno*. The superlative in Portuguese has the full paradigm of gender and number, and for languages with a richer nominal inflection, such as Russian, the number of inflectional forms can really explode because of this stacking.

Various types of morphological variants that are often considered inflectional are far from obligatory: the superlative only exists for gradable adjective, which is the minority of adjectives; and the female forms of nouns only exist for nouns denoting animate objects, which is an even smaller group.

Derived forms, being lemmas by themselves, can attract meanings of their own: the fact that a *movement* can be an ideological group is not related to it's being a derivation of *move*. But for inflections this should not hold. However, the word *cadeirinha* – which is the diminutive of *cadeira* (chair), can also mean a kind of saddle made of wood, or a *rickshaw* (in Macau). And the word *glasses* is no longer just the plural of *glass*.

3.2. The inflectional nature of derivation

There are several morphological processes that are almost completely inflectional in nature, but are still considered derivational, merely because they are category changing. For instance, various types of derivations are so productive that they are almost obligatory – almost any adjective has an adverbial form, and these adverbial forms are highly predictable as well: except for a small group of exceptions, adverbs in English are formed by putting *–ly* at the end of the adjective (with the appropriate phonological modifications); and in Dutch the adverbial form is simply identical to the adjective.

The nouns or adverbs derived from adjectives are hardly different from their original form in meaning – even more so if the derived noun is used attributively, or if the adjectival form applies to a deverbal noun: there is little difference between the sentences *he moved slowly* and *his movement was slow* accept from a change in perspective. In the case of the adjectival use of past participles, it is sometimes not even clear whether a particular use of the term is a past participle use, or an adjectival construction.

Given these considerations, Haspelmath (1995) argues to extend the class of inflections with word-class changing relations, or *transpositional inflection:* processes that are "*inflectional in the sense that it is regular, general, and productive, but nonetheless transpositional*" (Bauer, 2004). In the remainder of this article, *inherent inflection* will be used as possibly including transpositional inflections.

3.3. Paradigmatic Lexical Functions

The discussion above illustrates the fact that there is no clear boundary between inflection and derivation – the question whether a given relation should count as derivational or inflectional depends at least partly on the purpose of the question. But this duplicity creates a tension on their treatment in databases like MorDebe. On the one hand, for to purpose of consultation, it would be most useful if female nouns and superlatives would be modelled paradigmatically as part of their root. But since only lemmas can be related to meanings, they should be listed as separate lemmas.

The use of paradigmatic lexical function can very nicely relieve this tension. To take the female nouns as an example: the words *actor* and *actress* can be listed as separate lemmas, both with their own singular and plural forms. The two distinct lemmas are then explicitly linked by a lexical function (the lexical functions is taken from the DECIDE project: **female**). This explicit linking of the lemmas makes it possible to generate a larger paradigm for *actor* including both the male and female forms if the application requires it – the use of a functional link makes it possible to shift between an inflectional and a derivational perspective. The use of lexical functions in this fashion can also nicely explain for the stacking nature of inherent inflection because of the recursiveness of lexical functions.

In MorDebe, the set of relations that is considered strictly inflectional is largely reduced in this manner: only the number and gender of adjectives; tense, aspect, number and person for verbs; and number for nouns (despite its problems). All other modifications, some of which or traditionally taken as inflectional, are modelled as separate lemmas - these include female forms of nouns, superlatives, diminutives, but also deverbal nouns, deadjectival adverbs, etc.

In the Portuguese MorDebe database, all such inherent inflections are being modelled as derivational relations between different lemmas, where the relation between the lemmas is modelled by means of lexical functions. For practical purposes, the **s0** are separated into deadjectival nouns (**s0a**) and deverbal nouns (**s0v**); the LF's currently used are: **s0a, s0v, adv0,**

a2, able2, max, min, dim (diminutive)**, aum** (augmentative)**, female,** and **alt** (for orthographic variants: **alt**(*dourar*) = *doirar*). For Portuguese, about 15.000 LF relations have been recorded thus far, and all words that are added to the database are added together with their inflectional relations when applicable. This makes MorDebe one of the few large-scale LF databases in existence.

4. Lexical Functions over lemmas

Despite the appeal of the use of lexical functions as described in the previous section, there is a problem with it: the lexical functions are applied to *lemmas*, whereas LF's are by default defined over word-senses. Lexical functions are normally semantic derivations in at least three senses: firstly, they are not form based relations, since *"no morphological link needs to exist between the units involved"* (Polguère, 2000). Secondly, the relation they express is a semantic relation – even in the case of the so-called syntactic functions: **s0** expresses a nominalisation *with the same meaning* as its original. And thirdly, their arguments are word-senses rather than lemmas.

This last point is a direct consequence of the definition of a lexical function: lexical functions are defined as relations of *lexical units*, which in MTT are either *lexemes* or *phrasemes*, and a *lexeme* is *"a word taken in one well-specified sense and supplied with all the information fully determining its behavior when it is used in this sense"* (Mel'cuk, 1993, p. 250)[2]. The word-sense nature of LF's is easy to see: **Magn**(*hand*) = *dab*, but only in its meaning of *"someone with skill"*.

4.1. Inflectional Lexical Functions

An important difference between derivations and inflections is that inflections operate on the level of lemmas: the past tense *jumped* is not the past tense of a specific meaning of *jump*. Inflectional paradigms belong to the lemmas, and not to the word-senses. Because of this, the inflection-like derivations discussed in the previous section also have to be modelled over lemmas, and not over words senses. Hence, the lexical functions described in 3.3 range over lemmas - they are a different type of lexical functions, which will be called *inflectional functions*[3]. Given the fact that their arguments are entities of a different type, inflectional functions are strictly speaking not lexical functions at all, but different kinds of relations. This despite the fact that they are modelled in the same way as lexical functions, even use the same function names where available.

Inflectional functions are in a sense midway between morphological derivations and the semantic derivations used in MTT. On the one hand, they represent semantic relations, and not morphological processes. But on the other hand, the functions relate to lemmas, and not to their meanings. This is best illustrated by some examples.

In Portuguese, the derivational suffix *-mento* turns verb into nomina - modelled by the inflectional function **s0v**: **s0v**(*arrolhar*) = *arrolhamento* (corkage). But it is not the morphological derivation that is mapped by **s0v**, since the same function also maps different patterns: **s0v**(*canalizar*) = *canalização* (canalisation), **s0v**(*adastrar*) = *adastragem*

[2] In this article, the term *word-sense* will be used for sense-specific lexical units to avoid confusion with the sense-independent use of the term *lexeme*.
[3] The term *inflectional lexical function* was dropped on the suggestion of one of the reviewers.

(straightening), **s0v**(*assar*) = *assadura*, **s0v**(*abafar*) = *abafação* = *abafamento* = *abafadura* = *abafadela* (suffocation). The relation between the argument and value does not even have to be a productive derviational process: **s0v**(*propor*) = *proposição* (to propose), **s0v**(*fundir*) = *fundição* = *fusão* (foundation). In certain cases, there does not even have to be a morphological relation between the two: **max**(*bom*) = *ótimo* (good; best) - similar to irregular inflections: *went*, *mice*.

The other way around, not all relation which look derivational are modelled as such: although *dentadura* could be the **s0v** of *dentar* (to bite) - it is not. It is morphologically derived from *dentado* (toothed) and means 'set of teeth' or 'denture'. And even though *embocadura* (mouth of a river) is morphologically the nominalisation of *embocar* (to put to the mouth), it is not the **s0v** of it, because semantically, it is not the *action or effect of embocar*. In English, *ignorance* is no longer transparently related to *ignore* (Bauer, 2004).

But inflectional functions are not as semantic as the (normal) lexical functions. They have to be relations over the lemma and not its meaning: *abafação* (suffocation) can never be the **s0v** of *sufocar* (suffocate), despite the fact that *abafação* expresses the nominal content of *sufocar*. Although the relation does not have to be strictly derivational, there has to be a strong paradigmatic link between the value and the argument.

And inflectional functions have to be productive over the entire lemma, and not restricted to one of its senses. For instance, *embicadura* is not the **s0v** of *embicar*, since it is only the nominalisation in a very specific sense: 'heading into a port'. Although it should be said that the relation can be blocked by certain word-senses: **female**(*amarelo*) = *amarela,* even though this female form can only be used for animate reading of the word (pale person), in the same sense that the plural of *agua* (water) can only be used in count-noun readings. Inherent inflections are more likely to be partially defective than contextual inflections.

4.2. Asymmetry

Although the inflectional functions operate on lemmas, the derived term is clearly not a lemma: although the way of expressing the **s0a** for *agudo* (sharp; biting; acute; serious; keen; in-depth) is *agudo* rather independently of its meaning, the reverse is not true: although *promoção* can be the **s0v** of *promover*, it also has a number of other meanings, including *sales*. Only in a specific meaning of the word is the derived form the **s0v** form of the root - although it might of course happen to be the only meaning.

The derivational meaning does not even have to be a prominent meaning of the derivate: the **s0a** of *cavo* ($hollow_A$) is *cavidade*, but the meaning *hollowness* is only a marginal meaning of *cavidade:* it much more prominently means *cavity* (or $hollow_N$), which bears the same root as *cavo*, but is not derived from it – it is an adapted form of the latin word *cavitate* (*cavity*).

The fact that the derived form can have other meanings besides being the transparent derivation is not surprising, and in fact one of the reasons why inflectional functions are treated as derivations (in the sense of generating new words) in first place. But it does mean that inflectional functions are asymmetrical: they take lemmas as their arguments, but yield word-senses as their values. This has to be taken into account in the interpretation of inflectional functions: **female**(*pensador*) = *pensadora* expresses the fact that the word for referring to a female *pensador* (thinker) is *pensadora*, not that the word *pensadora* is identical to the female form of *pensador*.

5. Conclusion

The problem with the gradual distinction between derivation and inflection in morphological database can nicely be bridged by using lexical functions: because of the use of LF's, MorDebe can provide a rich set of paradigmatic information for every lemma in the database, including contextual inflections, but also inherent inflections and even category changing "inflections".

Lexical functions for this purpose (inflectional functions) are fundamentally different from normal lexical functions in that they range over lemmas rather than word-senses. However, for those inflectional functions that have an LF counterpart, there is a strong relation between the two: an inflectional function over a lemma implies that the corresponding LF holds for all related word-senses.

References

Bauer, Laurie. 2004. The function of word-formation and the inflection-derivation distinction. *In:* Aertsen, Hannay & Lyall (eds.) *Words and their Places. A Festschrift for J. Lachlan Mackenzie.* Amsterdam: Vrije Universiteit.

Booij, Geert. 1995. Inherent versus contextual inflection and the split morphology hypothesis. *In:* Booij & van Marle (eds.) *Yearbook of Morphology 1995*. Dordrecht: Kluwer.

Booij, Geert. *To appear*. Inflection and derivation. *In:* Brown (ed.) *Encyclopedia of Language and Linguistics*. Oxford: Elsevier.

Bybee, Geert. 1985. *Morphology*. Amsterdam: John Benjamins.

Haspelmath, Martin. 1995. Word-class-changing inflection and morphological theory. *In:* Booij & van Marle (eds.) *Yearbook of Morphology 1995*. Dordrecht: Kluwer.

Janssen, Maarten. 2002. *SIM*u*LLDA: a Multilingual Lexical Database Application using a Structured Interlingua*. PhD Thesis, Utrecht University.

Lieber, Rochelle, 1980. *On the organisation of the lexicon.* PhD Thesis, MIT.

Mel'cuk, Igor A. 1993. The Future of the Lexicon in Linguistic Description. *In:* Ik-Wan Lee (ed.) *Linguistics in the Morning Calm 3: Selected papers from SICOL-1992*. Korea: Seoul.

Perlmutter, David. 1988. The Split Morphology thesis: evidence from Yiddish. *In:* Hammond & Noonan (eds.) *Theoretical Morphology*. San Diego: Academic Press.

Polguère, Alain. 2000. Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. *In:* Heid et al. (eds.) *Proceedings of the Ninth EURALEX International Conference*. Stuttgart: IMS.

Schultink, Hendrik. 1962. *De morfologische valentie van het ongelede adjectief in mordern Nederlands*. Den Haag: van Goor & Zonen.