Proper Names in Slovene: Implications for defaults in inflectional morphology

Joseph Paul Stemberger (stemberg@interchange.ubc.ca)  
Department of Linguistics, E270-1866 Main Mall, UBC  
Vancouver, B.C. V6T 1Z1 Canada

Abstract

Inflectional systems generally contain default patterns, such as pl. –s in English, which are used in the absence of a reason to use a nondefault pattern. Contentious points include whether there can be more than one default pattern, whether proper names must follow the default pattern, and whether default patterns cover all words in the language. In Slovene, nouns are inflected for 6 cases and 3 numbers, and may be of 3 genders. Masculine gender appears to be the general default, showing the most widespread generalization, while feminine inflections generalize only to nouns that end in –a, including proper names. Female proper names that do not end in –a receive no overt inflections, with differences between surnames and personal names. Proper names do not have to follow default patterns, and may be ineligible for any inflectional rule at all. While standard linguistic and connectionist approaches can deal with the facts, the Marcus-Clahsen-Pinker approach finds the data challenging.

Introduction

In linguistic theories, sentences have traditionally been analyzed as strings of symbols (morphemes) that link meaning to sound. Information such as number (e.g. plural), case (e.g. locative) and gender (e.g. masculine, feminine, animate, etc.) is often encoded in an inflectional affix that attaches to a base and expresses a portion of the word’s meaning (such as the –s of the English plural form dogs and the –ed of the past-tense form walked); attachment to the base is accomplished via rules or constraints. Some of these affixes are “regular” and constitute default patterns that are used when there is no reason to use any other pattern, while other affixes are “irregular” and are used only in a lexically-idiomatic fashion (as with the –en of the English plural oxen, or the vowel change in the English past-tense form sang, cf. infinitive sing). Traditional linguistic theories allow multiple defaults of two types. First, there can be two or more classes of words, said to be of different genders; the classes can be defined semantically or may be fully or partially arbitrary (so that e.g. the “same” word may be masculine in one language but feminine in another). Second, the classes may be conditioned by phonological factors (e.g. whether the base has one syllable or two), so that different affixes are default in words with different phonological properties. Depending on the nature of the defaults, some bases may not be covered by any pattern, and so cannot be inflected to express a given piece of inflectional meaning; such forms are referred to as indeclinable, either in general or with respect to a given inflectional category. While languages with limited inflectional systems (such as English and German) might be limited to a single affix on nouns (e.g. plural), languages with richer inflectional systems may have many noun inflectional affixes, which may be quite different for different lexical classes; affixes are grouped into paradigms, and a typical word of a given class systematically takes all affixes for that class; it is not the case that each inflectional affix is assigned separately. Indeclinable nouns can be exceptional to the whole paradigm, or just to individual categories within it.

Over the past 10 years, one theoretical approach to language development has proposed restrictions, such that default affixes have a narrower range of properties than has traditionally been held in linguistic theory. Marcus et al. (1995) and Clahsen (1999) argued that there can be only one default affix (possibly allowing for some lexical classes) and that default affixes must be used for an odd assortment of nouns, including: foreign words, acronyms, nominalizations of closed-class lexical items, and proper names. This odd assortment of nouns, they maintain, reflect noncanonical words which do not have lexical entries in the usual sense and so cannot have any idiosyncrasies, including irregular morphology. They do not explain in detail what such words do have; presumably they are stored in a location separate from the main noun lexicon (so they cannot be subject to analogy with irregular common nouns). No arguments were put forward as to why defaults should have this narrowed set of properties, nor was there any attempt to explore the ramifications of these proposals for a wider set of human languages or for a wider set of noun inflections than plural.

Responses to these claims have centred around whether they are strictly accurate even for German and English plurals. Some (e.g. Penke & Krause, 2002) have argued that German feminine nouns ending in schwa (all of which take the plural affix –en) act like defaults in addition to the putative default –s that Marcus et al. argue for. Others have shown that German surnames do take the –s plural in an almost monolithic fashion, while personal names are quite variable, with particular personal names preferring different plural affixes (e.g., Hahn & Nakisa, 2000). Surnames can be treated as having a special default (-s), even if –s can also appear to a lesser degree for other classes of nouns.

In this paper, I will provide details about inflection in Slovene, a South Slavic language spoken in Slovenia and adjacent parts of Italy, Austria, Hungary, and Croatia. Slovene has a much richer noun inflectional system than English or German, with six cases (nominative, genitive, dative, locative, accusative, and instrumental), three numbers (singular, dual, and plural), and three genders.
(masculine, feminine, & neuter), organized paradigmatically into four declensions (one masculine, two feminine, and one neuter, with several subpatterns and lexical idiosyncrasies). The masculine gender and associated declension generalizes to most words, except for words that end in /a/ in the nominative singular, to which one of the feminine declensions (and feminine gender) generalizes in a uniform fashion. Proper names are interesting, because names that refer to female human beings but do not end in /a/ receive no inflections at all; and surnames and personal names do not behave in a parallel fashion. Even feminine surnames that end in /a/ (and so can be inflected for case) show some behaviour that is quite distinct from masculine surnames.

These data have interesting implications for theories of human language. Standard linguistic theories are able to deal with the facts in a straightforward fashion, and connectionist models are also likely to be able to deal with them. The Marcus-Clahsen-Pinker approach, in contrast, has difficulty with the facts at a basic level. It is not clear that there exists a variant of their theory in which the facts can be dealt with, without a radical re-evaluation of the way that proper names are treated. Proper names are not automatically subject to default rules, there can be special defaults for proper names, and the full set of default rules can in fact fail to cover some proper names at all.

This paper uses standard linguistic methodology to explore the facts of Slovene. The patterns reported here are discussed in the standard reference grammars of Slovene (Toporišič, 2000; Herrity, 2000). Doleschal (2000) has demonstrated these patterns experimentally. The patterns are also true of casual speech, and are easily observable in printed materials such as books, magazines, and newspapers. Here I focus on names and on case inflections, which, unlike plural, are entirely natural and common even for personal names; while names such as Mary and Joe are rarely used in the plural, there is nothing unnatural about e.g. Dative case (for Mary, to Joe), and indeed case-inflected forms for names make up a sizable percentage of tokens of all words in Slovene, including proper names.

The genders and declensions of Slovene

Slovene has three genders making up four declensions. The singular suffixes are listed in the following table. Gender is additionally marked via suffixes on adjectives and on predicates. Where two suffixes are listed, the choice is conditioned by the phonological characteristics of the base (which is beyond the scope of this paper).

<table>
<thead>
<tr>
<th>SINGULAR</th>
<th>Masc</th>
<th>Fem (reg)</th>
<th>Fem (irreg)</th>
<th>Neut</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>---</td>
<td>-a</td>
<td>---</td>
<td>-o/-e</td>
</tr>
<tr>
<td>Genitive</td>
<td>-a</td>
<td>-e</td>
<td>-i</td>
<td>-a</td>
</tr>
<tr>
<td>Dative</td>
<td>-u</td>
<td>-i</td>
<td>-i</td>
<td>-u</td>
</tr>
<tr>
<td>Locative</td>
<td>-u</td>
<td>-i</td>
<td>-i</td>
<td>-u</td>
</tr>
<tr>
<td>Accusative</td>
<td>-a</td>
<td>-o</td>
<td>---</td>
<td>-o/-e</td>
</tr>
<tr>
<td>Instrumental</td>
<td>-om/-em</td>
<td>-o</td>
<td>-jol/-ijo</td>
<td>-om/-em</td>
</tr>
</tbody>
</table>

Default status

Masculine gender has the highest type and token frequency, and generalizes most broadly. The regular feminine declension in –a is also very frequent, and generalizes automatically to words ending in /a/. Neuter gender is least common in type and token frequency, occurring with only a few hundred monomorphemic nouns, and does not generalize to new words. Masculine gender appears to be the general default, but feminine –a is a secondary default.

MASCULINE NOUNS

Almost all native masculine nouns end in a consonant in the nominative singular form. Masculine gender and suffixes occur with the odd assortment of words that Marcus et al. (1995) proposed as a sign of default status, including the following (with the locative suffix highlighted):

(1) common nouns
   nos: o nosu ‘about the nose’

(2) personal names
   Jože: pri Jožetu ‘at Joe’s’

(3) surnames
   Štemberger: pri Štembergerju ‘at Stemberger’s’

(4) nominalized uses of closed class lexical items
   ampak ‘but (conj.)’: o takšnem ampaku ‘about this sort of but’

(5) acronyms
   RTV (Radiotelevizija): zakon o RTV-ju ‘the law about RTV’ (/erteveju/)

(6) (unassimilated) foreign words
   na mojem computerju ‘on my computer’

That this is not simply generalization to similar words is demonstrated by the fact that masculine gender is generalized to most vowel-final forms, while very few masculine native common nouns end in vowels. Acronyms (5 above) and loanwords that end in any vowel other than /a/ receive masculine gender. For all vowels other than /o/, a /j/ is inserted between the base and the affix. This is true for common nouns (7), masculine personal names (8), and masculine surnames (9), even though these words do not closely resemble any native words.

(7) tabu: na tamkajšnjem tabuju ‘in the tabu there’

(8) pri Toni ‘at Toni’s’

(9) pri Mussoliniju ‘at Mussolini’s’

For native proper names that end in /o/, there are two observed patterns. In the standard pattern, the /o/ disappears in all inflected forms, and the affixes are added as usual.
(10), but in colloquial speech, the /o/ is often retained, and a /t/ is inserted between the /o/ and the affix. Loanwords that end in /o/ always lose the /o/ (11).

(10) Vilko: pri Vilku 'at Vilko's' pri Vilkotu

(11) avto: pri avtu 'by the car'

FEMININE NOUNS
Slovene has two declensional classes for feminine common nouns. There is a small set of feminine forms that end in a consonant; this is not a productive class, and no loanwords are ever assimilated into this class.

(12) miš: pri miši 'at the mouse's house'

There is a very large set of nouns that end in –a and are a member of this class. Many names belong to this class, as do assimilated and unassimilated foreign words and words that are (or were originally) acronyms, all signs of default status according to Marcus et al. (1995).

(13) common nouns
mačka: o mački 'about the cat'

(14) personal names
Marija: pri Mariji 'at Mary's'

(15) surnames
Iskra: pri Mariji Iskri 'at Mary Iskra's house'

(16) loanwords
pica: na pisci 'on the pizza'

(17) unassimilated foreign names
Joanna: pri Joanni 'at Joanna's'
Florida: na Floridi 'in Florida'

(18) acronyms pronounced as a word (rather than string of letters) note: accusative case
a. velik uspeh za Naso
   'big success for NASA'
b. Cheney hotel izjemo za Cio
   'Cheney wanted an exception for the CIA.'

NEUTER NOUNS
Neuter nouns end either in /o/ or /e/ (with the vowel conditioned by the preceding consonant). Note that neuter nouns occur as place names, with no tendency to change to masculine gender. Case endings in the singular are identical with the masculine; neuters are primarily distinguished by the plural nom./acc. suffix –a (cf. masc. –i/-e) and by neuter agreement with adjectival and verbal forms.

(19) common noun
jabolko 'apple': jabolka (pl.)

(20) place names
Novo mesto
   ('new town')

Neuter gender is not productive (Dressler & Makovec-Černe, 1995). As shown in (11) above, loanwords that end in /o/ are not borrowed with neuter gender, but take default masculine. (This is itself an interesting fact with implications for cognitive models of morphology, but it is beyond the scope of this paper.) That place names based on common nouns such as mesto 'town' retain neuter gender, even though neuter gender is not generalized to loanwords, suggests that the Marcus-Clahsen-Pinker view of proper names is at best an oversimplification.

Grammatical gender & natural gender
There are many nouns that refer to non-gendered things in the world and which arbitrarily take masculine or feminine gender. However, the labels “masculine” and “feminine” are used because human referents with natural gender (such as ‘man’ and ‘woman’, as well as ‘Frank’ and ‘Mary’) show the appropriate grammatical gender. Such words generally occur in two forms: a masculine form (which may be monomorphic or derived) and a feminine form (which often, but not always, contains a derivational affix):

(21)a. 'neighbour': sosed (masc.), soseda (fem.)
b. 'friend': prijatelj (masc.), prijateljica (fem.)
c. 'journalist': novinar (masc.), novinarka (fem.)

Nonetheless, there are some words that appear in only one form and do not respect natural gender:

(22)a. oseba (fem.) 'person' (male or female)
   b. otrok (masc.) 'child' (male or female)

In addition, there are a few words that end in –a that can refer to both men and women, and that take masculine agreement with men and feminine agreement with women, even when the inflected forms are overtly feminine in form:

(23)a. masc.: dober vođa 'good leader' (nom.)
   dobrevođe 'of a good leader' (gen.)
   b. fem.: dobra vođa 'good leader' (nom.)
   dobrevođe 'of a good leader' (gen.)

The almost-obligatory association of grammatical gender with the natural gender may possibly play a role in accounting for the inflection of feminine proper names in some theories.

WHAT ARE THE DEFAULTS?
Standard analyses would assume two default patterns: (1) masculine gender and associated suffixes, which generalize to consonant-final forms (which resemble many masculine nouns) and to vowel-final forms (which do not); (2) feminine gender, which generalizes to all forms that end in /a/ in the nominative singular (though a small number of masculine nouns also take feminine suffixes). The masculine declension appears to be a general default, while
feminine is restricted to nouns that end in /a/. Neuter gender is not a default, and does not generalize to loan words that end in /a/.

The feminine declension in –a acts like a default. There is no reason to treat it as unproductive or irregular in any way, despite being restricted to a specific phonological environment: words that end in /a/ in the nominative singular. The Marcus-Clahsen-Pinker approach would recognize two genders, each with default affixation rules, where feminine (but not masculine) gender is restricted to a particular region in phonetic space. Why neuter –o is nonproductive is less clear; the small number of words involved should not be relevant to the Marcus-Clahsen-Pinker approach or to standard linguistic approaches. Only simulations will reveal whether connectionist models can account for the lack of productivity for neuter –o.

Lack of inflection on feminine proper names

Not all feminine proper names are inflected, however. All feminine proper names that end in any phoneme other than /a/ are indeclinable (i.e., appear with a single form in all uses, regardless of case).

There is a class of feminine personal names that end in /i/, which cannot be overtly declined (24), but adjectives and verb forms do take feminine gender via agreement (25).

(24) Mici: za Mici ‘for Mici’ (*za Mico/Micija)
(25) Mici je prišla. ‘Mici arrived (fem.).’
(cf. masc. Jože je prišel. ‘Joe arrived.’)

Although these hypocoristics (nicknames) are old enough to be thought of as “native”, the pattern was most likely borrowed from German. These names occur in all syntactic positions, but without overt case marking. Unassimilated foreign feminine personal names that do not end in /a/ follow this same pattern:

(26)a. pri Bridget ‘at Bridget’s’ (*pri Bridgeti)
b. pri Sue ‘at Sue’s’ (*pri Sueji)
(27) Bridget je prišla. ‘Bridget arrived (fem.).’

The situation is very different for feminine surnames that do not end in /a/. Feminine surnames rarely occur independently in a sentence. Whereas uninflected personal names are acceptable and common in Slovene sentences, uninflected feminine surnames are not possible by themselves. Uninflected surnames are nonetheless common as part of a larger construction, such as personal-name+surname or title+surname.

(28)a. za Britney Spears ‘for Britney Spears’
b. pri Pamela Anderson ‘at Pamela Anderson’s house’
c. Gospo Stemberger ‘for Mrs. Stemberger’

To explain the contrast between the free occurrence of uninflected feminine surnames in these larger constructions, vs. the absence of such surnames by themselves, Doleschal (2000) proposed that the surname occupies an adjunct position in the syntax, such as is commonly observed with phrases such as ‘the Union Hotel’, where the word ‘hotel’ is inflected, but the name of the hotel can optionally remain uninflected:

(29) v Hotelu Union ~ v Hotelu Unioni
‘in the Union Hotel’

One difference is that uninflected feminine surnames obligatorily go into this adjunct position, whereas other nouns can optionally be in the adjunct position (and hence uninflected) or in a regular syntactic position (and hence in the appropriate case). Interestingly, feminine surnames that end in /a/, and which can be inflected, vary between inflected and uninflected forms in this position:

(30) pri Mariji Iskri ~ pri Marij Iskra
‘at Marija Iskra’s house’

While indeclinable surnames are forced into the adjunct position, any feminine surname can go into that position optionally, like common nouns in constructions such as Hotel Union.

In contrast, a masculine surname cannot appear uninflected even in these larger constructions:

(31) za Kevina Costnerja (*za Kevina Costner)
‘for Kevin Costner’

Masculine surnames are excluded even with surnames such as Janša that end in /a/ and take feminine case suffixes:

(32)a. z Janezom Janšo ~ z Janezom Janšem
b. * z Janezom Janša

It is not the case that surnames in general are allowed in this adjunct position, but feminine surnames are: obligatorily when they have no overt case marking, optionally even when they can show overt case marking. The syntax clearly treats masculine and feminine surnames in a non-parallel fashion.

USE OF DERIVED STEMS FOR NAMES

It is possible for feminine surnames to appear in any syntactic position, but they require a derivational suffix before they can receive case inflections. The derivational suffix –ov- (with the variants –ev- and –jev- in predictable phonological environments) can be added, with appropriate suffixes, and the name can then appear in any syntactic position:

(33)a. Spearsova ni noseča. ‘Spears is not pregnant.’
b. pri Spearsovi ‘at Spears’s house’
c. za Spearsovo ‘for Spears’

Such usage is statistically uncommon, accounting for less than 10% of uses of feminine surnames.
This derivational suffix (historically a form of the possessive adjective) can also appear optionally for plurals:

(34) Štembergerjevi ~ Štembergerji ‘the Stembergers’

Such plurals contradict the Marcus-Clahsen-Pinker claim that plurals of surnames must strictly follow default patterns for the language. While the derivational suffix is neither needed nor available for singular masculine surnames, it constitutes the sole mechanism for feminine surnames that do not end in /a/ to receive case marking and thus to appear in independent case-marked syntactic positions.

The problem of the indeclinable feminine names

Slovene is a highly inflected language, and forms which systematically receive no overt inflections are startling. That feminine suffixes such as nom. –a and loc. –i do not generalize to foreign names such as Bridget and to native surnames such as Žajc reinforces the conclusion based on productivity above, that the regular feminine declension is restricted to one region of phonetic space, with /a/ at the end of the base. (This presupposes that the final /a/ is actually part of the base, rather than a suffix marking nominative singular, as most linguistic theories would assume; the details need to be worked out for this new sort of analysis.)

A more interesting question is why such words do not appear with default masculine suffixes. If the masculine declension supplies default suffixes for common nouns that do not end in /a/, why shouldn’t it do so for names referring to females? The Marcus-Clahsen-Pinker view of the default as an emergency measure for supplying affixes predicts that the (masculine) defaults should appear with these words, especially since their approach requires that proper names must in principle show default inflectional patterns. Why is it that non-/a/-final proper names referring to female humans do not take any default affxes at all?

Solution 1: Feminine natural gender blocks grammatical masculine gender. A single noun cannot be both masculine and feminine at the same time. Whether a noun is masculine vs. feminine clearly will depend on the nature of the procedures that assign gender. The key observation here is that Slovene almost always respects the natural gender of humans, leading to a large number of pairs of words, one set applied to males and the other applied to females. It is reasonable to assume a procedure that assigns feminine grammatical gender to words that refer to female humans, including names. This would conflict with (and take precedence over) any procedure that would assign (default) masculine gender to nouns. Thus, female personal names and surnames such as Bridget and Štemberger cannot take masculine gender and hence cannot take the suffixes of the associated masculine declension. However, since they do not end in /a/, they also cannot take the suffixes of the regular feminine declension. These names fall through the cracks: they in principle cannot be subject to either set of default inflectional patterns in the language. They surface without overt inflectional suffixes.

The Marcus-Clahsen-Pinker approach incurs a cost with this analysis: a modification of the view that default affxes are last-resort emergency ways to create inflected forms. There would instead be circumstances in which even these “emergency” measures fail. The result is a form without overt inflectional suffixes. But names such as Mici and Bridget can be used in any syntactic position, suggesting that they are in fact treated as forms that carry case and gender (even though they show no overt marking of case or gender). The forms thus “have” e.g. locative case in some sense, but have what is known as “zero-marking” in the acquisition and adult processing literatures (e.g. Stemberger & MacWhinney, 1986). While this is a part of some approaches (e.g., Stemberger, 2002), Marcus et al. (1992) make clear that it is not compatible with their approach. This solution requires a significant modification of the Marcus-Clahsen-Pinker approach: in the horse-race between defaults and irregulars, the uninflected base is always one of the competitors and can win under some conditions. This modification calls into question the places in which Pinker and Prince (1988) criticize the Rumelhart and McClelland (1986) connectionist model for failing to output overtly inflected forms under some circumstances. Until we have an empirically derived list of circumstances under which default rules can fail, no model can be criticized for failing under some circumstances.

Interestingly, for feminine surnames, even a zero-marked form is unavailable. Instead, the name must receive a derivational suffix first. This implies that derivational and inflectional affxes are part of the same competition. This can be stated in some linguistic approaches (e.g. Optimality Theory: McCarthy & Prince, 1993), and is likely to be learnable in connectionist models on the basis of input, but is not predicted by the Marcus-Clahsen-Pinker approach.

Solution 2: Indeclinable feminine proper names as special inflection classes. We can take zero-marked (indeclinable) feminine personal names as a special inflectional class, where zero is the default for this semantically-defined class. Because these personal names are inflected for case (though not overtly so), they can appear freely in all syntactic positions where personal names can appear; as feminine nouns, they trigger feminine agreement under the appropriate syntactic circumstances. A second class contains surnames, which are not assigned overt case and so are restricted to syntactic adjunct positions within noun phrases, unless they undergo a derivational rule (-ov-) which allows them to take inflections. Note that neither class can be fully semantically defined, because proper names that end in -a are not members of either class. Pinker and Prince (1988) severely criticized connectionist models for outputting forms without overt inflectional endings, with phonological conditioning. Positing two classes, defined partly in terms of semantics and partly in terms of form, is most definitely not in the spirit of their approach. Nor is positing special inflectional classes for proper names, within a theory that requires proper names to take default inflectional patterns.
Conclusion

Slovene proper names present a complex set of data that is challenging for some cognitive models of morphology. In an attempt to differentiate their models from connectionist models, Pinker and Prince (1988), Marcus et al. (1992, 1995), Clahsen (1999), and Pinker and Ullman (2002), among others, have proposed some very specific properties that differentiate default morphological patterns from irregular patterns. One key point is a restriction on proper names, such that they are required to take default affixation. Slovene feminine names that do not end in /a/ are problematic for this approach, because they are simply undeclined (zero-marked), with no affixes at all, default or irregular. On the surface, the Marcus-Clahsen-Pinker approach does not appear to be able to handle the facts.

The only workable solutions are (a) to posit one or two inflectional classes that are specifically made up of feminine proper nouns that do not end in -a, with different classes for personal names and for surnames, or (b) to posit that default rules may fail to apply to some forms, and that the true emergency measure for inflections is zero-marking (i.e. no overt inflection). The surface patterns can probably be learned on the basis of the frequency of undeclined feminine surnames, and of the frequency of hypocoristics ending in /i/ (such as Miči). While such analyses are available to standard linguistic approaches that put few restrictions on proper names, and quite plausibly could be learned by statistically-driven models such as connectionist models, they are not conceptually compatible with the Marcus-Clahsen-Pinker approach to inflection. With either solution, there is no need to posit that proper names are not listed in the lexicon: they simply constitute a special class of nouns, along with count vs. mass nouns, masculine vs. feminine vs. neuter nouns, animate vs. inanimate nouns, etc.

Proper names, as well as defaults, do not have some of the characteristics posited in the Marcus-Clahsen-Pinker approach, which consequently means that proper names do not have the implications for models of cognition that they claim.

Acknowledgments

This research was supported by an individual research grant from NSERC. I would like to thank Vilko Maček for helpful discussions about the facts of Slovene.

References
