

Numerical Forms in Welsh-language Software

Introduction

This document discusses the issues for Welsh translators and developers which arise with software which generates user interface messages containing numbers, such as "5 files found". It describes typical software design patterns, and how they interact with Welsh initial consonant mutation rules. It then recommends best practice for

1 Localizable Strings

Popular software is commonly available with more than one interface language. The interface text is generally authored in a source language (often English), and then translated to multiple target languages via a process of localization (L10N). The L10N process typically consists of making available to a translator the source-language strings of the software, which are then individually translated into the target language.

2 String Templates

In some cases the strings are simple text (as in example 1.1 below). In other cases, the string acts as a template for any number of possible strings; for example, with the information in example 1.2 below, the software can generate the English strings "Found message from Hicin", "Found message from Siencin", "Found message from Siac" etc., or their Welsh equivalents as necessary. The required name is inserted into the string in place of a 'placeholder' string (the %s in this case).

Fig. 1: Example string templates in English and Welsh

Example	English String	Welsh String
1.1	Print Document	Argraffu Dogfen
1.2	Found message from %s	Wedi canfod neges o %s
1.3	Time remaining: %d minutes	Amser ar ôl: %d munud
1.4	Page %d of %d	Tudalen %d o %d

3 Grammatical Number

A particular issue arises when inserting numbers into string templates (in which case the placeholder will sometimes be %d, with "d" being a mnemonic for "decimal"), because different languages commonly have different rules for agreement between numbers and nouns/units. For example:

- English has a singular form for $n = 1$ ("1 house") and a plural form otherwise ("0 houses", "2 houses", "3 houses" etc).
- French has a singular form for $n = 0$ or 1 ("0 maison", "1 maison") and a plural form otherwise ("2 maisons", "3 maisons", etc).

Numerical Forms in Welsh-language Software

- Polish has a singular form for $n = 1$ ("1 dom"), a 'paucal' form for $n = 2, 3, 4, 22, 23, 24, 32, 33, 34$ etc ("2 domy", "3 domy", "4 domy", "22 domy" etc), and a plural form otherwise ("5 domów", "6 domów" etc).

This makes it infeasible in many languages to create a single template from which the different number forms can be generated. The internationally standard way of handling this is that the software is informed which number form categories are required for a particular target language (See Fig. 2 below). **Number form categories may be defined once, globally, for each software development framework. They cannot always be changed on a per-string or even per-application basis.** This limitation is important to the discussion below.

Subsequently, when translating the string templates, the translator must provide one translated string template for each number form category in that language. This provides greater grammatical correctness, at a cost of increased complexity and workload.

Fig. 2: Number form categories for different languages

Language	Categories
English	singular ($n = 1$) plural (otherwise)
French	singular ($n = 0$ or 1) plural (otherwise)
Polish	singular ($n = 1$) paucal ($2 \leq n \bmod 10 \leq 4$ and $n \bmod 100$ not between 10 and 20) plural (otherwise)

4 Agreement in Welsh counting

Grammatical number agreement is fairly simple in Welsh counting, as the singular form is always used ("1 afal", "2 afal", "3 afal", "4 afal" etc), except with the number zero which takes the plural form ("0 afalau"). There is also the paraphrasis 'o + plural form' which can be used for two or more items, but has slightly different connotations.

There is, however, an additional issue in Welsh not related to grammatical number, in that different numbers cause different patterns of initial consonant mutation (see **Fig. 3**)

Numerical Forms in Welsh-language Software

Fig. 3: Initial consonant mutations caused by Welsh cardinal numbers

Masculine Example	Feminine Example
0 cŵn	0 cathod
1 ci	1 gath (soft mutation)
2 gi (soft mutation)	2 gath (soft mutation)
3 chi (aspirate mutation)	3 cath
4 ci	4 cath
5 ci	5 cath
6 chi (aspirate mutation)	6 chath (aspirate mutation)
7, 8, 9, 10, 11, 12, 13 ... ci	7, 8, 9, 10, 11, 12, 13 ... cath

This is just a special case of a more general issue with Welsh string templates, in that substituted forms may cause or undergo mutation. A Welsh string template such as that in Fig. 4 is likely to result in grammatically incorrect substituted strings.

Fig. 4: Example of a grammatical error in a substituted string

<i>English template:</i>	This is the train for %s
<i>Welsh template:</i>	Dyma'r trê'n i %s
<i>Welsh placename:</i>	Bangor
<i>Gramatically incorrect substituted string:</i>	*Dyma'r trê'n i Bangor

5 Using Grammatical Number Rules for Welsh Mutations

It is in principle possible to define grammatical number categories for Welsh so that initial consonant mutations will be handled correctly, as in Fig 5 below.

Fig 5: Grammatical number categories for Welsh mutations

Numbers	Notes
0	No mutation; plural noun
1	Soft mutation for feminine only
2	Soft mutation
3	Aspirate mutation for masculine only
6	Aspirate mutation
Other (including fractional)	No mutation

Numerical Forms in Welsh-language Software

6 Drawbacks of complex number rules

The rules in Fig. 5 are sufficient to represent all words correctly other than a few exceptions (e.g. "diwrnod" day) which follow different mutation patterns. However there are significant disadvantages. Defining these rules globally for a particular software development framework may in some cases mean that Welsh translators have to supply six translated string templates for every source template which contains a number placeholder, as in Fig. 6 below.

Fig. 6: templates following grammatical number rules in Welsh and English:

Language	Form	Text
English	Singular	%d document found
English	Plural	%d documents found
Welsh	Zero	Wedi canfod %d dogfen
Welsh	Singular	Wedi canfod %d ddogfen
Welsh	Dual	Wedi canfod %d ddogfen
Welsh	Triple	Wedi canfod %d dogfen
Welsh	Sextuple	Wedi canfod %d dogfen
Welsh	Other	Wedi canfod %d dogfen

There will always be multiple identical templates (as there are six templates and only three mutations). Furthermore it is likely to be unclear to many translators how to input these forms, as the number categories required (0, 1, 2, 3, 6, other) do not neatly fit Welsh grammatical concepts (such as Radical, Soft Mutation and Aspirate Mutation).

Note that the situation worsens where a string contains two numerical placeholders (such as "%d pages in %d chapters"), as the number of possible combinations multiply to 36. Note also that the additional complexity only deals with numerical forms, and cannot guarantee correctness in other cases such as Fig 4 above.

Some software development frameworks may shield the translator from some of the complexity described above, e.g. by allowing the input of a single template form to cover all possibilities where this is linguistically acceptable.

7 Alternative Solutions in Use

Welsh is not particularly unusual in exhibiting grammatical issues with string substitution; the same is also true in many other morphologically rich languages where, for example, word endings in the template may need to be varied depending on the grammatical category of the substituted word. Amongst speakers of such languages there is usually a degree of tolerance to such shortcomings. Indeed, English speakers readily accept strings such as those in Fig. 7 below.

Numerical Forms in Welsh-language Software

Fig. 7: Ambiguous singular/plural English strings seen in published application software.

5 file(s) found
1 notification(s)
You have 0 message(s)

In cases where the sophistication and complexity described above cannot be justified, or is not available, Welsh strings with numerical placeholders are generally translated with a single form, as in Fig. 8 below.

Fig. 8: Single Welsh template and example generated string.

Wedi canfod %d dogfen
Wedi canfod 0 dogfen
Wedi canfod 1 dogfen
Wedi canfod 2 dogfen
Wedi canfod 3 dogfen
Wedi canfod 4 dogfen

Such generated strings are generally accepted by users of Welsh-language software; although the mutation pattern is not strictly correct, the resulting pattern is arguably more natural to read than the English examples using hybrid singular/plural forms such as "file(s)".

This solution is currently used by most popular systems today, including Firefox, Microsoft Windows + Office and OpenOffice.org, as well as websites such as Google and Facebook.

8 Recommended Best Practice

1. Where the complexity is justified and can be handled, correct mutation patterns for cardinal numbers can be achieved by using six grammatical number categories in Welsh localization, namely 0, 1, 2, 3, 6 and other (see Fig 5 above).
2. Otherwise, a single grammatical category should be used for all numbers, using the unmutated form of the noun in translation (see Fig 8 above).
3. Care must be taken when defining Welsh rules for a software development framework, so as not to force translators to follow the six-category pattern if this may be inappropriate.

David Chan
Language Technologies Unit
Bangor University