

A Short Description of Icelandic Compounds

Kristín Bjarnadóttir
Orðabók Háskólans

April 23, 2002

1 Introduction

Literature on Icelandic compounding is not easily available, even in Icelandic. The following description was concocted in connection with a review of some NLP material; it is a rough version of the general rules of compounding in Icelandic.

The rules are however fairly useless when it comes to individual compounds, as not all possibilities are acceptable, even if grammatically correct. The choice of form seems to be lexicalized, as shown in the examples in the article *Lexicalization and the Selection of Compounds for a Bilingual Icelandic Dictionary Base* (KB 1995).¹

2 Rules on Icelandic Compounding

1. Compounds can be formed from words belonging to all major categories, and all these categories can be both left- or right-attached. The productivity varies greatly between categories, nouns and adjectives being the most productive, both on the left and right of the compounds.
2. Rules for compounding are recursive and some rules of derivation and compounding are mutually recursive. (The interaction of affixation and compounding is a topic I will not go into; it is a veritable jungle.) The productivity of recursive rules varies, according to the category of the formal head of the word.
3. The word forms used on the left in compounds can be without ending (stem), or with inflectional endings. Up to four different inflectional forms are theoretically possible for each noun. Link phonemes do also occur, as do sound changes in stems, producing combining forms not occurring in

¹The paper also contains some comments on homophonous forms which are a great problem in Icelandic compounds, and of course also in any kind of automatic syntactic analysis.

isolation (bound forms). Some nouns have up to five different forms on the left in compounds, but commonly there are three variants. The choice between variants is arbitrary but not free. Some inflectional classes pose restraints on the choice of their combining forms.

These are the possible forms for nouns as 1st part of compounds:

(a) Stems

- i. Same form of stem as in the nominative singular (basic dictionary form):

*orð+myndun*² ‘word formation’ (*orð* neut. + *myndun* fem.)

hest+hús ‘horse house’, i.e. ‘stable’ (*hestur* masc. + *hús* neut.)

The word form in dictionary entries is (usually) the nominative singular. The nominative singular ending disappears in the compounds, as in *hest.ur*.

- ii. Stems without umlaut where umlaut is found in the basic form of the lexeme:

nagl+rót ‘base of finger-/toe-nail’ (*nögl* ‘nail’ fem.nom.sg. + *rót* ‘root’ fem.)

nas+hyrningur ‘rhinoceros’ (*nös* ‘nose’ fem.nom.sg. + *-hyrningur* ‘horned one’ masc. (bound form))

- iii. Stems with umlaut (rare):

lög+brot, *rök+leysa* (plural stems)

- iv. Irregular sound changes in the stem:

mann+tal ‘census’ (*mann-* from *maður* ‘man’ nom.sg. + *tal* ‘count’ neut.)

(b) Inflected word forms as first part of compound:

- i. Genitive singular:

borðs+horn ‘corner of a table/desk’ (gen.sg. of *borð* ‘table’ fem. + *horn* ‘corner’ neut.)

bókar+kápa ‘dustjacket’ (gen.sg. of *bók* ‘book’ fem. + *kápa* ‘coat’, ‘cover’ fem.)

penna+strik ‘stroke of a pen, line’ (gen.sg. of *penni* masc. + *strik* ‘stroke’ neut.)

- ii. Genitive plural:

orða+bók ‘dictionary’ (gen.pl. of *orð* ‘word’ neut. + *bók* ‘book’ fem.)

²For clarity, ‘+’ is added between parts of compounds in the examples.

- iii. Datives, singular and plural: This is very rare and the constraints are very rigid. A list of 282 words of this type can be found on the Institute of Lexicography's website, <http://www.lexis.hi.is/datsams.html>.
- (c) Link phonemes: Cf. section 3 below.
4. Other categories in compounds:
- (a) Nouns can be on the left of compounds belonging to all word classes. The same rules apply as in the noun+noun compounds, but the distribution of the different forms is more rigid.
- i. noun+adjective:
- barn+góður* 'good to children' (*barn* = stem)
biblíu+fróður 'bible' + 'learned' (*biblíu-* = gen.sg.)
- ii. noun+verb:
- hand+þvo* 'wash by hand' (*hand* = stem without umlaut)
mal+bika 'lay tarmac' (*mal* = stem without umlaut)
greindar+prófa 'test sby's intelligence' (*greindar-* = gen.sg.)
gufu+þvo 'wash with steam' (*gufu-* = gen.sg.)
- iii. noun+adverb:
- blindhliðar+megin* 'on (your) blind side' (*blindhliðar-* gen.sg.)
- (b) Adjectives on the left of compounds:
- i. Stem:
- blá+ber* 'blueberry' (adj. + noun)
ill+menni 'bad man' (adj. + noun)
fagur+eygur 'with beautiful eyes' (adj. + adj.)
fín+kornóttur 'finely grained' (adj. + adj.)
fast+binda 'determine' (sth.; lit. 'to set sth. fast') (adj. + verb)
fín+mala 'grind finely' (adj. + verb)
- ii. Inflected forms, genitive:
- lítills+verður* 'worth little' (*lítills* = gen.sg.) (adj. + adj.)
ills+vitandi 'bad knowing' (i.e. conscious of badness) (*ills-* = gen.sg.) (adj. + adj.)
sjúkra+hús 'sick house' (i.e. hospital) (*sjúkra-* = gen.pl.) (adj. + noun)
lítills+virða 'count sby. as of little worth' (adj. + verb)
- iii. Inflected forms, neuter:
- heitt+elskaður* 'dearly beloved' (*heitt* = neut. of *heitur*) (adj. + adj.)
hátt+virtur 'highly estimated' (*hátt* = neut. of *hár*) (adj. + adj.)

hæst+virtur ‘most highly estimated’ (*hæst* = neut.superlative of *hár*) (adj. + adj.)

iv. Inflection inside compound:

The first part of some compounds has the definite adjectival form which inflects for case inside the compound:

Nom. *litlifingur*, acc. *litlafingur*

The comparative and superlative can also appear as the first part of compounds:

Nom. *hæsti+réttur* ‘highest court’ (i.e. ‘court of appeals’), acc. *hæstarétt*; the combining form of the compound is the genitive: *hæstaréttar+dómari* ‘highest court judge’
nom. *efri+gómur* ‘upper gum’

(c) Verbs can appear on the left of compounds, often with link phonemes as in *brenni+merkja* ‘brand’ v., *tyggi+gúmmí* ‘chewing gum’ neut. Stems rarely occur in this position, but cannot be ruled out. The rules for this kind of compound formation seem to be quite restrained.

(d) Adverbials, NPs, PPs and other such phrasal items also occur on the left of compounds, as in *utanbæjarmaður* ‘man from outside town’, and *milliþinga+nefnd* ‘between sessions committee’. In other words, Scalise’s No Phrase Constraint (1986) does not apply.

5. Any category can be right-attached in compounds. Here are a few more examples of right attachment of nouns and adjectives. The right-attachment of verbs and adverbials is also possible. (Similar examples are also found above.)

(a) Nouns:

blá+móða ‘blue mist’ (adj. + noun)

bindi+kraftur ‘binding force’ (verb + noun)

(b) Adjectives:

stór+ríkur ‘big rich’, i.e. ‘very rich’ (adj. + adj.)

bjart+eygur ‘bright-eyed’ (adj. + adj.)

fleyti+fullur ‘floating full’, i.e. ‘full to the brim’ (verb + adj.)

6. Multiple (recursive) compounding:

The choice of combining forms on the left in multiple compounding is not as free as in compounds formed from two base words. The trend is towards genitives in the first part of these compounds, as in *skrifborðs+plata* (gen.sg.) vs. *borð+plata* (stem). This is a very strong trend, not a firm rule. (Cf. Baldur Jónsson 1984.)

3 Link phonemes

Link phonemes in Icelandic are exceptions to the general rules on compounding in nouns, and could as such be listed. It would not be sufficient to list the combining forms containing link phonemes as the same lexeme could also appear in regularly formed compounds, either as a stem or as a genitive form. Link phonemes also occur with other word classes, as seen in the following examples. The link phonemes are not common enough to pose a major problem in NLP projects on par with the differences in form in the compounds.

1. Link phonemes in noun + noun compounds:

-a- *elda+buska*
-i- *fiski+bolla*
-s- *kúplings+diskur, glettnis+bros, hæðnis+tónn*
-u- *ráðu+nautur*

2. Link phonemes in combining verbal forms:

-i- in *ali+fugl* ‘domestic fowl’
(*ala* verb (‘breed’) + *fugl* fem. (‘bird’))
brennimerkja ‘brand’
brenna verb (‘burn’) + *merkja* verb (‘mark’)

3. Link phonemes also occur in affixation, as with *-legur* when added to verbs. The status of the affix *-legur* is in itself somewhat dubious, as it adds on to inflected word forms, which is otherwise a feature of compounding, not affixation.

-(a)n- *aðdán+legur, heyran+legur, greinan+legur*

4 Combining forms

There are numerous word forms only occurring inside compounds, either as first or last part. In the database of approximately 60 thousand compounds compiled for my MA-thesis there are over six hundred of these. The three main types of bound forms are as follows:

1. Cranberry morphemes:

Classic cases of berries, such as *hind-* in *hind+ber* ‘raspberry’, etc. There is a multitude of these forms, but they cannot be said to be a part of productive word formation.

2. Variants of stems:

An example is *nátt-* as a variant of *nótt* ‘night’, in common compounds such as *nátt+föt* ‘night clothes’, i.e. ‘pyjamas’. The regular combining

form would be the genitive, *natur-*, as in *natur+vakt* ‘night shift’ or *natur+líf* ‘night life’.

3. Derived bound forms:

Very many Icelandic nouns have right-attached bound variants, such as the derived forms with the affix *-i*. Examples: *-hveli*, *-hýsi*, *-yrði*, *-menni* from *hvalur* ‘whale’, *hús* ‘house’, *orð* ‘word’, and *maður* ‘man’. There are a number of affixes involved.

References

- Baldur Jónsson. 1984. Samsett nafnorð með samsetta liði. Fáeinar athugasemdir. *Festskrift til Einar Lundeby*. 3. október 1984, p. 158–174.
- Kristín Bjarnadóttir. 1995. Lexicalization and the Selection of Compounds for a Bilingual Icelandic Dictionary Base. Ásta Svavarsdóttir, Guðrún Kvaran, Jón Hilmar Jónsson (ed.). *Nordiske studier i leksikografi* 3:255-263.
- Kristín Bjarnadóttir. 1996. *Afleiðsla og samsetning í generatífri málfræði og greining á íslenskum gögnum*. [Derivation and Compounding in Generative Grammar: Analyzing Icelandic Data.] MA-thesis. Unpublished manuscript, University of Iceland.
- Kristín Bjarnadóttir. 2000. *Þágufallssamsetningar í ritmálssafni Orðabókar Háskólans*. [Dative Compounds in the Institute of Lexicography’s Written Language Archive.]. <http://www.lexis.hi.is/datsams.html>.
- Scalise, Sergio. 1986. *Generative Morphology*. Foris, Dordrecht.