

Heikki Särkkä

IMPROVING THE USABILITY OF THE FINNISH COMPREHENSION DICTIONARY

A problem encountered by anyone trying to use a bilingual Finnish dictionary for the purpose of comprehending a Finnish text is that it is not always easy to equate a given word form with the appropriate dictionary form. This is due to the fact that Finnish has an exceptionally rich morphology. Traditional dictionaries reduce this wealth to nominative singular for nouns and the infinitive for verbs. A natural consequence is that very few texts can be comprehended with the help of a dictionary only. Normally, of course, a person consulting a bilingual dictionary will know the language to some extent. On the whole, the use of a dictionary for decoding an unknown language is only possible when the so-called unknown language is in fact closely related to a language one knows. Examples of language pairs for which such a method works could be Spanish and Italian, Russian and Ukrainian, and Finnish and Estonian.

Under certain conditions, however, it is possible to translate from a language one does not know. To do this, the translator must know the subject matter. Second, he must be equipped with the proper tools over and above conventional dictionaries. One example of such a tool is the JAPANESE-ENGLISH GRAMMAR DICTIONARY by J. Jelínek, which makes it possible for a person with no knowledge of the Japanese language to translate not only individual words, but whole sentences. Traditional Japanese-English dictionaries are, of course, needed as well, but they would be insufficient on their own. To give a somewhat simplified description of the grammar dictionary, it consists basically of tables of alphabetically arranged grammatical elements which allow a step-wise left-to-right decoding of the word and of the sentence.

A full-scale Finnish-English grammar dictionary on the same lines would be perfectly feasible. In what follows I am setting myself the more modest task of examining what features of such a grammar dictionary could be incorporated into comprehension dictionaries of Finnish to make them more serviceable to non-native users.

Above I referred to the complexity of the morphological component of Finnish grammar. In what follows, special attention will be paid to only one aspect of the topic, the morphology of the noun. Maximally, the noun has the following structure:

STEM + NUMBER + CASE + (POSSESSIVE) + (ENCLITIC)

Using the above formula, Karlsson (1982) calculates that the noun has at least 2,112 different forms. The aim of the present paper is to outline a manual lemmatization system that allows the dictionary user to find the nominative singular of any Finnish noun.

The JAPANESE-ENGLISH GRAMMAR DICTIONARY rests on the principle

of left-to-right decoding of the word. This is possible because, with certain exceptions, the stem of a noun, verb, or adjective is written in kanji, the Chinese ideograms, while the rest of the word, the grammatical elements, are written in the hiragana syllabary. As there is no comparable way of knowing where the stem of a Finnish noun ends, it seems preferable to start the decoding from the right-hand end of the word, where the choice of possible elements is smallest.

Although textually the most frequent structural analysis for a noun is STEM + NUMBER + CASE, we have to be prepared for noun forms that include the possessive and/or the enclitic. Appendix 1 shows the different 'enclitic particles' occurring in Standard Finnish. The use of the capital indicates a context-dependent choice between a front and a back vowel, so that A stands for a/ä, O for o/ö, and U for u/y.

Once the enclitics have been stripped off, the word may have a 'possessive' suffix (Appendix 2).

Before dealing with the structure STEM + NUMBER + CASE, I shall briefly discuss the phenomenon of 'gradation'. Basically, gradation is a phonological process by which the stops /p t k/ are weakened in a closed syllable. The consonant(s) occurring at the syllable boundary immediately preceding an open syllable are said to represent the strong grade (SG), while those occurring at the beginning of a closed syllable are said to represent the weak grade (WG). The different forms are shown in Appendix 3. The centrality of gradation in Finnish phonology makes it a constant source of difficulty in finding the lemma of an inflected noun.

Perhaps because existing bilingual Finnish dictionaries are primarily intended for Finnish users, gradation has been given insufficient treatment from the point of view of the non-native user. As a concession to them, nouns subject to gradation have been indicated as such by an asterisk. A useful further aid would be a table showing the various alterations caused by gradation. At present, no dictionary gives this information.

After this digression, I shall come back to the lemmatization of nouns of the structure STEM + NUMBER + CASE. As an example I shall take the genitive singular. The genitive singular invariably has the ending -n. Stripping it off, we are left with the genitive singular stem (henceforth to be called the 'genitive stem' for short). There are a dozen or so nouns that have an irregular genitive. The list of exceptional genitives has to be consulted first. If the genitive stem under consideration is not found on that, it is taken to the Genitive Conversion Table (Appendix 4). This shows the regular correspondences between genitive stems on the one hand and nominative stems on the other. It can still happen that the particular genitive stem we are considering does not fit any of the descriptions given in the conversion table. It should then be assumed that the genitive stem is identical with the nominative stem with the exception of having a weak grade consonant. To find out whether this is the case, the table entitled Gradation (Appendix 3) may be used. Finally, if the genitive stem does not have a consonant that is subject to gradation, the nominative stem is identical with the

genitive stem, e.g. talo 'house', gen. sg. talon.

A similar analysis can be carried out of all other case suffixes, or, to put it more precisely, of all other strings identical with a case suffix. Theoretically, it is possible to examine each structure position as a separate problem, identifying, for example, the suffix occurring in position CASE as a partitive suffix and the one in position NUMBER as a sign of the plural. However, in view of the fact that there is no suffix identifying a noun as a singular form, another procedure seems preferable. First, all strings having the structural description NUMBER + CASE are identified. There are some fifty of them altogether. After that, each is provided with its particular description: partitive singular, partitive plural etc. Where possible, contextual conditions would be given for each string under which it would be possible for it to have a certain grammatical function. For example, the string -ita can be a partitive plural suffix only when preceded immediately by a vowel. To each string identifiable as a particular case suffix would be appended a translation equivalent, or, as the case may be, several of them.

Traditionally, the *raison d'être* of most bilingual dictionaries with Finnish as the source language has been to serve as an aid in producing foreign language texts. In consequence, their usefulness as comprehension dictionaries of Finnish has been limited. Yet progress has been made in the past ten years. Katara and Schellbach-Kopra's FINNISCH-DEUTSCHES GROSSWÖRTERBUCH gives outline paradigms for the 85 noun declension classes identified by Nykysuomen sanakirja. It also gives the declension class of each noun. The FINSKO-RUSSKIJ SLOVAR' by Vahros and Sherbakoff does the same, but also often gives the genitive singular where the genitive and nominative stems are different. In spite of these improvements, existing Finnish-foreign language dictionaries are still primarily production dictionaries. If they are to be made more effective as comprehension aids, they should be provided with tables listing all suffixes that occur in a given structural position. Such tables would not turn them into complete grammar dictionaries, but would greatly add to their value as comprehension tools for non-Finnish users.

#### Reference

Karlsson, F. (1982) Suomen kielen äänne- ja muotorakenne. Porvoo: WSOY

APPENDIX 1

ENCLITIC PARTICLES

Particle	Function or meaning
hAn	Signals shared knowledge
kAAAn	neither, nor
kin	also, even, too
kO	Signals question
pA	Signals surprise

APPENDIX 2

POSSESSIVE SUFFIXES

Left context	Person	Sg	Pl
vowel	1	ni	mme
vowel	2	si	nne
vowel	3	nsA	
A		An	
e		en	

APPENDIX 3

GRADATION

	Nominative sg	Genitive sg	SG / WG
1.	lakki 'cap'	laki/n	kk / k
2.	tippa 'drop'	tipa/n	pp / p
3.	hytti 'cabin'	hyti/n	tt / t
4.	vako 'furrow'	vao/n	k / Ø
5.	tapa 'habit'	tava/n	p / v
6.	sota 'war'	soda/n	t / d
7.	lanka 'thread'	langa/n	nk / ng
8.	kampa 'comb'	kamma/n	mp / mm
9.	ranta 'shore'	ranna/n	nt / nn
10.	ilta 'evening'	illa/n	lt / ll
11.	parta 'beard'	parra/n	rt / rr
12.	jälki 'trace'	jälje/n	lki/lje
13.	järki 'reason'	järje/n	rki/rje
14.	pohje 'calf (of leg)'	pohkee/n	hk / hj
15.	virka 'office'	vira/n	rk / r
16.	jalka 'leg'	jala/n	lk / l
17.	uhka 'threat'	uha/n	hk / h

	NUMBER OF GRADABLE WORDS	NUMBER OF OCCURRENCES IN THE CORPUS	PER CENT OF ALL CASES OF GRADATION IN THE CORPUS
t: d	80	14474	29
tt: t	95	13483	26
k: Ø	25	6162	12
nt: nn	38	6026	12
kk: k	22	4441	9
rt: rr	9	1544	3
p: v	7	1390	3
pp: p	8	924	2
nk: ng	6	932	2
lt: ll	8	730	.1
mp: mm	5	534	1
lk: lj	3	320	.6
k: v	1	267	.5
<b>Total</b>	<b>307</b>	<b>51227</b>	<b>100</b>

APPENDIX 4

GENITIVE CONVERSION TABLE

GENITIVE SINGULAR

N:B: Where the consonant(s) to the left of each line represent the strong grade (SG), the weak grade (WG) must be chosen on the left of the nominative stem. Conversely, if the consonant(s) to the left of the line do not represent the strong grade, the indication of weak grade on the left of the nominative stem should be ignored.

Capital vowel letters stand for either member of the vowel harmony pair, that is, A stands for a/ä, O for o/ö and U for u/y.

Genitive stem	Nominative stem	Examples	CONT.
(SG) + AA	(WG) + As	altaa- / allas 'pool'	
(SG) + Are	(WG) + Ar	sisare- / sisar 'sister'	
de	1) s	rakkaude- / rakkaus 'love'	
	2) si	uude- / uusi 'new'	
(SG) + ee	1) (WG) + e	kokee- / koe 'test'	
	2) Ut	tullee- / tullut	PCPL
(WG) + e	1) (SG) + i	kive- / kivi 'stone'	
	2) ø	askele- / askel 'step'	
hde	1) hti	lahde- / lahti 'bay'	
	2) ksi	kahde- / kaksi 'two'	
(SG) ii	(WG) + is	alttii- / altis 'eager'	
kse	s	teräkse- / teräs 'steel'	
x+V+x+ immA	x+V+x+ in	kovimma- / kovin	SUP
x+V+x+ AmmA	x+V+x+ Ampi	vakavaumma- / vakavampi	COMP
x+V+x+ emmA	x+V+x+ empi	kovemma- / kovempi	COMP
x+V+x+ OmmA	x+V+x+ Ompi	isomma- / isompi	COMP
x+V+x+ UmmA	x+V+x+ Umpi	rajumma- / rajumpi	COMP
nne	nsi	kanne- / kansi 'lid'	
rre	rsi	hirre- / hirsi 'log'	
V + se	V + nen	punaise- / punainen 'red'	
Ue	Ut	ohue- / ohut 'thin'	
UvU	UkU	puvu- / puku 'suit'	
ttoma	ton	puuttoma- / puuton 'treeless'	
(SG) + ime	(WG)+ in	hapettime- / hapetin 'oxidizer'	

If none of the above apply, change WG into SG.