

Arto Anttila

## Prosodic Constraints on /-ntV/ in Finnish

### Abstract

The distribution of Finnish nominalizers /-nti/, /-ntO/, and /-ntA/ is constrained by word prosody, in particular syllable weight. We describe the distribution of these morphemes and point out some of their theoretical implications.

### 1. Preliminaries

In Finnish, word prosody plays an important role in shaping morphological patterns.<sup>1</sup> The central notion is syllable prominence, which refers to at least three different dimensions of the syllable: stress, weight, and the sonority of the nuclear vowel (see e.g. Anttila 1997a). Consider the word *apteekkari* ‘pharmacist’:

(1) Three dimensions of syllable prominence

	Prominent	Unprominent
Stress	Stressed (‘σ)	Unstressed (σ)
	<i>á</i> p.teek.kà.ri	<i>á</i> p.teek.kà. <i>r</i> i
Weight	Heavy (H)	Light (L)
	<i>á</i> p.teek.kà.ri	<i>á</i> p.teek.kà. <i>r</i> i
Vowel sonority	Low vowel (A)	High vowel (I)
	<i>á</i> p.teek.kà.ri	<i>á</i> p.teek.kà. <i>r</i> i

The first syllable *áp* is prominent on all three counts: it is stressed, heavy, and has a low-vowel nucleus.<sup>2</sup> The last syllable *ri* is unprominent on all three counts: it is unstressed, light, and has a high-vowel nucleus. The

---

<sup>1</sup> An earlier version of this article appeared as part of Anttila and Revithiadou 2000. I thank an anonymous reviewer for helpful suggestions.

<sup>2</sup> In Finnish, (C)V-syllables are light. All other syllables, e.g. (C)VC, (C)VV, (C)VVC, (C)VCC, are heavy (but see Keyser & Kiparsky 1984). As usual, syllable onsets are irrelevant for weight.

medial syllables *teek* and *kà* have mixed properties: *teek* is heavy, but unstressed and has a mid-vowel nucleus; *kà* is light, but (arguably) stressed and has a low-vowel nucleus.

Word prosody is built around the notion of rhythm, i.e. alternating prominences, which emerges particularly clearly in the domain of stress (Lieberman and Prince 1977, Prince 1983). In Finnish non-compound words, primary stress falls on the initial syllable and secondary stress falls on every other syllable after that, subject to some interesting complications (Sadaniemi 1949, Carlson 1978, F. Karlsson 1983: 150–155, Hanson & Kiparsky 1996, Elenbaas & Kager 1999). In other words, the stress rule scrupulously avoids creating sequences of adjacent stressed syllables (clashes) and adjacent unstressed syllables (lapses), respecting the rhythmic principle. There is also evidence for the avoidance of adjacent heavy syllables and adjacent light syllables which emerges quantitatively in morphological variation (Itkonen 1957, Anttila 1997b). In what follows, we note that the distribution of the nominalizer morphemes /-nti/, /-ntO/, and /-ntA/ provides further evidence for the rhythmic principle in morphology.

## 2. The prosodic patterning of /-ntV/

The distribution of /-nti/, /-ntO/, and /-ntA/ is easy to state in prosodic terms. The examples in (2)–(4) illustrate the basic pattern.

- (2) /-nti/ after heavy syllables only
- |                            |                 |                 |
|----------------------------|-----------------|-----------------|
| <u>juo</u> -nti            | ‘drink-nom’     | ‘drinking’      |
| si. <u>jai</u> -nti        | ‘locate-nom’    | ‘location’      |
| ar.vi. <u>oi</u> -nti      | ‘estimate-nom’  | ‘estimation’    |
| for.ma.li. <u>soi</u> -nti | ‘formalize-nom’ | ‘formalization’ |
- (3) /-nta/ after light syllables only
- |                           |                  |               |
|---------------------------|------------------|---------------|
| lu. <u>e</u> -nta         | ‘read-nom’       | ‘reading’     |
| las. <u>ke</u> -nta       | ‘count-nom’      | ‘counting’    |
| pa.hek. <u>su</u> -nta    | ‘disapprove-nom’ | ‘disapproval’ |
| ve.te.leh. <u>di</u> -ntä | ‘loiter-nom’     | ‘loitering’   |
- (4) /-nto/ after both heavies and lights
- |                    |               |             |
|--------------------|---------------|-------------|
| <u>luo</u> -nto    | ‘create-nom’  | ‘nature’    |
| <u>pyy</u> -ntö    | ‘request-nom’ | ‘request’   |
| a. <u>su</u> -nto  | ‘inhabit-nom’ | ‘apartment’ |
| us. <u>ko</u> -nto | ‘believe-nom’ | ‘religion’  |

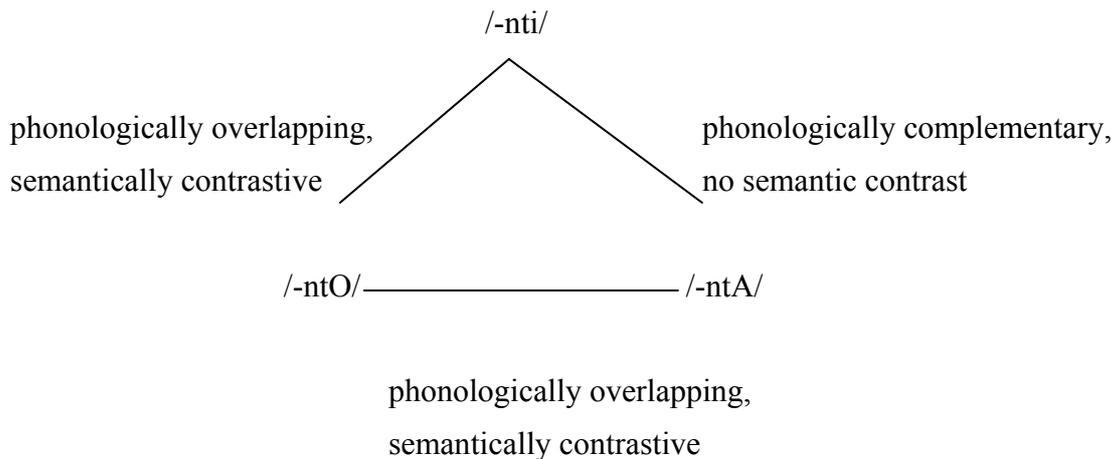
The above data illustrate two facts. First, /-nti/ and /-ntA/ are prosodically in complementary distribution: /-nti/ attaches to heavy syllables and /-ntA/ attaches to light syllables. Second, /-nti/ and /-ntA/ are semantically transparent and do not differ in meaning. /-ntO/ is different on both counts: it appears to have no prosodic restrictions and it is often associated with unpredictable meanings. Indeed, many verbs have a semantically transparent /-nti, -ntA/ nominalization as well as a lexicalized /-ntO/-nominalization. The following doublets are typical:

(5)	luo-nti	‘creating’	luo-nto	‘nature, creation’
	lue-nta	‘reading’	lue-nto	‘lecture’
	laske-nta	‘counting’	laske-nto	‘elementary arithmetic’
	halli-nta	‘governing’	halli-nto	‘government’
	istu-nta	‘sitting’	istu-nto	‘session’
	kuki-nta	‘flowering’	kuki-nto	‘blossom’
	palki-nta	‘rewarding’	palki-nto	‘prize’
	polje-nta	‘treading’	polje-nto	‘beat’
	tutki-nta	‘examining’	tutki-nto	‘exam’

An anonymous reviewer points out that words in the first column can often be replaced by non-finite *-minen* forms without any appreciable change in meaning, whereas those in the second column appear more idiosyncratic and more ‘nouny’.

The distribution of /-ntV/-morphemes can be summarized as in (6):

(6) The distribution of the three nominalizer morphemes



We now look at the distribution in more detail. Table (7) classifies all the /ntV/-final non-compound nominals in *Suomen kielen käänteissanakirja* (KSK) (Tuomi 1972) based on the weight and stress of the last syllable of the stem with the final /ntV/ removed: 'H = stressed heavy, H = unstressed heavy, 'L = stressed light, L = unstressed light. Both derived and non-derived lexemes are included. Primary and secondary stresses are not distinguished, but an example of each is given if available. If an /ntV/-nominalization exists, an example is given in boldface.

(7) /ntV/-final nominals in KSK (Tuomi 1972)

	'H	H	'L	L	Total:
/-nti/	96.8% [596] <b>vie-n.ti,</b> <b>vá.ki.òì-n.ti</b>	14.3% [1] <b>sí.jai-n.ti</b>	9.1% [7] tún.ti	1.1% [7] sé.kun.ti	[611]
/-ntO/	2.6% [16] <b>lúo-n.to</b>	71.4% [5] <b>há.vai-n.to</b>	24.7% [19] kún.to, és.pe.ràn.to	12.3% [78] <b>tút.ki-n.to</b>	[118]
/-ntA/	0.6% [4] <b>sáa-n.ta</b>	14.3% [1] <b>án.sai-n.ta</b>	66.2% [51] hín.ta, <b>lúon.neh.dì-n.ta</b>	86.6% [548] <b>tút.ki-n.ta</b>	[604]
Total:	100% [616]	100% [7]	100% [77]	100% [633]	[1,333]

The weight-based generalization is virtually exceptionless. The suffix /-nti/ is almost exclusively limited to the context after stressed heavies; it is marginally found after unstressed heavies, the only case being *sijai-nti* 'location'; and it is never found after lights. The suffix /-ntA/ is almost exclusively limited to the context after light syllables; it is marginally possible after heavy unstressed syllables, the only case being *ansai-nta* 'earning'; and it is next to nonexistent after stressed heavies: the forms *saa-nta*, *tee-ntä*, and *syö-ntä* listed in KSK do not seem part of current usage. No clear prosodic pattern emerges with /-ntO/.

We conclude that the descriptive statement “/-nti/ after heavy syllables, /-ntA/ after light syllables, /-ntO/ anywhere” is essentially correct and reference to stress is unnecessary. However, the data in (7) also reveal a subtler pattern: the proportion of final /nti/ steadily decreases on the scale 'H > H > 'L > L, whereas /ntA/ shows the opposite tendency. This suggests that the selection of /-ntV/ is a special case of a more general prominence dissimilation pattern that holds quantitatively within the entire lexicon.

### 3. Analysis

The basic generalization is that the high-prominence nucleus /A/ in /-ntA/ seeks out low-prominence contexts and the low-prominence nucleus /i/ in /-nti/ seeks out high-prominence contexts. The selection of /-ntV/ thus seems a special case of the rhythmic principle that favors alternating prominences. The problem is that the generalization only holds for the three nominalizer morphemes. A general constraint like \*L.I ‘Avoid light syllable + high vowel sequences’ would rule out the ungrammatical derived forms \**luonnehdi-nti* and \**tutki-nti*, but it would also rule out the grammatical nonderived forms *tunti* ‘hour’, *synti* ‘sin’, *sekunti* ‘second’, and *Islanti* ‘Iceland’. The latter sound phonologically impeccable and speakers do not feel compelled to convert them to \**tunta*, \**syntä*, \**sekunta*, and \**Islanta* respectively. Similarly, a general constraint like \*H.A ‘Avoid heavy syllable + low vowel sequences’ would rule out the ungrammatical derived forms \**vie-ntä* and \**sijai-nta*, but it would also rule out the grammatical nonderived form *suunta*. Again, speakers do not feel compelled to convert the latter to \**suunti*. One might take this as evidence that the phonological generalization is illusory and that suffix selection is plainly morphological. However, this would not be satisfactory either because it would not provide any rationale for the nearly perfect phonological distribution of /-nti/ and /-ntA/.

We propose the following analysis. The lexicon contains verb roots as well as the suffixes /-nti/, /-ntO/, /-ntA/. Any verb root can be freely combined with any suffix and the ill-formed combinations are ruled out by ranked and violable constraints in the sense of Optimality Theory (Prince & Smolensky 1993/2004). For describing the basic pattern, we need the four constraints in (8a), ranked as in (8b). A sample evaluation is shown in (9).

- |     |                |   |
|-----|----------------|---|
| (8) | a. FAITH(ROOT) | Realize the features [±high ±low] of a root morpheme. |
|     | FAITH          | Realize the features [±high ±low] of a morpheme.      |
|     | *H.A           | Avoid sequences of heavy syllables and low vowels.    |
|     | *L.I           | Avoid sequences of light syllables and high vowels.   |

b. FAITH(ROOT) >> {\*H.A, \*L.I} >> FAITH

(9) **Sample tableaux:** A light syllable followed by a /-ntV/ suffix

		FAITH(ROOT)	*H.A	*L.I	FAITH
/laske-nta/	laske-nti			*!	**
	laske-nto				*!
	→ laske-nta				
/laske-nto/	laske-nti			*!	*
	→ laske-nto				
	laske-nta				*!
/laske-nti/	laske-nti			*!	
	→ laske-nto				*
	laske-nta				**!

This grammar predicts both *laske-nta* (L-nta) and *laske-nto* (L-nto), but rules out *\*laske-nti* (L-nti) due to the ranking  $*L.I \gg FAITH$ . The potential input /laske-nti/ is neutralized to *laske-nto*. The surface form *laske-nto* is thus ambiguous as it can be derived from two distinct inputs: /laske-nto/ and /laske-nti/. By Lexicon Optimization (Prince and Smolensky 1993/2004), which is essentially a restatement of Postal's Naturalness Condition (Postal 1968: 53–77), the input phonetically closest to the output, i.e. /laske-nto/, is the actual input.<sup>3</sup>

The special behavior of nonderived roots can be captured by the constraint FAITH(ROOT) which is a special case of FAITH (McCarthy and Prince 1995). The observation that roots tend to be more resilient under phonological pressure than affixes is a familiar one. As shown in (10), the proposed grammar predicts no neutralization in roots. All three inputs surface faithfully due to the high-ranking FAITH(ROOT). The analysis works analogously for H.A sequences.

<sup>3</sup> The complete candidate set contains many candidates not discussed here for reasons of space. In particular, the candidate suffixes /-nte/, /-ntu/, and /-nty/ must be ruled out by some constraints insofar as their absence is due to systematic phonological reasons.

(10) **Sample tableaux:** A light syllable followed by /ntV/ within a root

		FAITH(ROOT)	*H.A	*L.I	FAITH
/islanti/	→ islanti			*	
	islanto	*!			*
	islanta	*!*			**
/esperanto/	esperanti	*!		*	*
	→ esperanto				
	esperanta	*!			*
/hinta/	hinti	*!*		*	**
	hinto	*!			*
	→ hinta				

Why has only /-ntO/ developed specialized meanings? The answer is probably related to the fact that only /-ntO/ is able to make it through the phonology unneutralized. The same is true of nonderived roots although for a different reason. In contrast, /-nti/ and /-ntA/ are phonologically more vulnerable and subject to neutralization depending on the prosodic environment. The general pattern seems to be that semantic distinctions enter the language through environments where phonological markedness constraints are weak or irrelevant and phonological contrasts are consequently able to surface. A parallel example is stem-final /A/-deletion, e.g. *kano-j-a* ‘hen-PL-PAR’ from /kana/ (no deletion) vs. *mun-i-a* ‘egg-PL-PAR’ from /muna/ (deletion). This alternation is largely phonologically conditioned, but exhibits morpholexical conditions under strictly limited phonological circumstances, namely in trisyllabic stems where the left context contains a non-high non-round vowel followed by a coronal consonant, e.g. *kiharo-i-ta* ‘curl-PL-PAR, noun’ vs. *kihar-i-a* ‘curly-PL-PAR, adjective’ (G. Karlsson 1978). For an analysis, see Anttila 2002.

#### 4. Going further

The analysis sketched above describes the distribution of /-ntV/-suffixes, but it is too specific to be of much interest in the long run. We have already seen evidence that the real pattern may be more general: the proportion of final /nti/ steadily decreases on the scale 'H > H > 'L > L, whereas /ntA/ shows the opposite pattern. This suggests two hypotheses: (i) both weight and stress contribute to syllable prominence, but stress plays a subsidiary role; (ii) the rhythmic principle is reflected quantitatively in the

phonological structure of the entire lexicon, implying that the same syllable prominence constraints play a role in categorical alternations as well as in quantitative phonotactics.

It would be a straightforward matter to test these hypotheses systematically in a large data set such as *KSK*. Here we simply note that both hypotheses hold up if we expand the data set slightly. The data in (11) represent a pseudo-random 2,578 word sample of the Finnish lexicon, more precisely the vowel-final nominal lexical entries from A to F in *KSK*, including nonderived stems (Anttila and Revithiadou 2000). This is approximately 6% of the stems in the dictionary. This time, we divide the left context into three categories of weight: SH = superheavy syllable (*pa.tent.ti* ‘patent’, three moras), H = heavy syllable (*sa.met.ti* ‘velvet’, two moras), and L = light syllable (*pa.pe.ri* ‘paper’, one mora). We emphasize that these figures are preliminary and further empirical work is needed to test the generalization.

(11) The last two syllables of V-final stems in *KSK*

	'SH	SH	'H	H	'L	L
/i, u, y/ %	85	80	75	71	49	30
/o, ö/ %	2	8	11	13	15	25
/a, ä/ %	13	12	14	16	36	45
	100	100	100	100	100	100

These facts invite us to develop a unified theory of categorical alternations and quantitative phonotactics. The suffixal /-ntV/-alternations in (2)–(4) and the quantitative lexical patterns in (11) are clearly related and should follow from the same phonological principles. Spelling out these principles and their interactions remains a challenge for future research.

## References

- Anttila, Arto (1997a) *Variation in Finnish Phonology and Morphology*. Ph.D. thesis, Stanford University.
- (1997b) Deriving variation from grammar. In Frans Hinskens, Roeland van Hout & W. Leo Wetzels (eds.) *Variation, Change, and Phonological Theory*, pp. 35–68. Current Issues in Linguistic Theory 146. Amsterdam & Philadelphia, PA: John Benjamins.
- (2002) Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20: 1–42.

- Anttila, Arto & Anthi Revithiadou (2000) Variation in allomorph selection. In Masako Hirotsu, Andries Coetzee, Nancy Hall & Ji-yung Kim (eds.) *Proceedings of the North East Linguistic Society 30, Volume 1*, pp. 29–42. Amherst, MA: Graduate Linguistics Student Association.
- Carlson, Lauri (1978) Word stress in Finnish. Unpublished manuscript, Massachusetts Institute of Technology.
- Elenbaas, Nine & René Kager (1999) Ternary rhythm and the lapse constraint. *Phonology* 16: 273–329.
- Hanson, Kristin & Paul Kiparsky (1996) A parametric theory of poetic meter. *Language* 72.2: 287–335.
- Itkonen, Terho (1957) Mellakoihin vai mellakkoihin? Yleiskielemme eräiden taivutushorjuvuuksien taustaa. *Virittäjä* 61: 259–286.
- Karlsson, Fred (1983) *Suomen kielen äänne- ja muotorakenne*. Porvoo, Helsinki & Juva: Werner Söderström Osakeyhtiö.
- Karlsson, Göran (1978) Kolmi- ja useampitavuisten nominivartaloiden loppu-A:n edustuminen monikon *i:n* edellä. In Alho Alhoniemi, Jussi Kallio, Mauno Koski, Päivi Rintala & Kalevi Wiik (eds.) *Rakenteita. Juhlakirja Osmo Ikolan 60-vuotispäiväksi, 6.2.1978*, pp. 86–99. Turku: Turun yliopisto.
- Keyser, Samuel J. & Paul Kiparsky (1984) Syllable structure in Finnish phonology. In Mark Aronoff & Richard T. Oehrle (eds.) *Language Sound Structure*, pp. 7–31. Cambridge, MA & London: The MIT Press.
- Liberman, Mark & Alan S. Prince (1977) On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249–336.
- McCarthy, John J. & Alan S. Prince (1995) Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, pp. 249–384. University of Massachusetts Occasional Papers 18. Amherst, MA: Graduate Linguistics Student Association.
- Postal, Paul M. (1968) *Aspects of Phonological Theory*. New York: Harper & Row.
- Prince, Alan S. (1983) Relating to the grid. *Linguistic Inquiry* 14:19-100.
- Prince, Alan & Paul Smolensky (1993/2004) *Optimality Theory: Constraint interaction in Generative Grammar*. Malden, MA: Blackwell Publishing.
- Sadeniemi, Matti (1949) *Metriikkamme perusteet ja sovellusta moderneihin ja antiikin mittoihin*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Tuomi, Tuomo (1972) *Suomen kielen käänteissanakirja. Reverse Dictionary of Modern Standard Finnish*. [Helsinki]: Suomalaisen Kirjallisuuden Seura.

Contact information:

Arto Anttila  
 Department of Linguistics  
 Stanford University  
 Stanford, CA 94305, U.S.A.  
 Anttila(at)Stanford(dot)edu  
<http://www.stanford.edu/~anttila/>