LSA.365 | Variation in Optimality Theory, Handout #2
Arto Anttila

WORD STRESS IN FINNISH

Arto Anttila
Stanford University

1. Word Stress in Finnish

(1) Literature on Finnish word stress: Carlson 1978, Elenbaas 1999, Elenbaas and Kager 1999, Hanson and Kiparsky 1996, Karttunen 2006, Karvonen 2005, Keyser and Kiparsky 1984, Kiparsky 2003, Sadeniemi 1949.

(2) Primary stress is always initial. All the action is in secondary stress.

(3) The general pattern (binarity): Stress every other syllable after primary stress.

(a)	ká.las.tè.let	'you are fishing'
(b)	ká.las.tè.le.mì.nen	'fishing'
(c)	íl.moit.tàu.tu.mì.nen	'registering'
(d)	ú.jos.tè.le.màt.to.mùu.des.tàn.sa	'from his lack of shyness'

(4) The special pattern (ternarity): Skip a light syllable if the syllable after that is heavy unless the heavy syllable is final in which case skipping is optional.

(a)	ká.las.te.lèm.me	'we are fishing'
(b)	íl.moit.tàu.tu.mi.sès.ta	'from registering'
(c)	vói.mis.te.lùt.te.le.màs.ta	'from causing to do gymnastics'
(d)	rá.vin.tò.lat ~ rá.vin.to.làt	'restaurants'

(5) Analysis (first approximation): Binary trochaic feet from left to right; main stress on the leftmost foot; heavy syllables must be stressed.

(6) Problems:
- (a) Plenty of variation and apparent exceptions to the secondary stress rules.
- (b) Secondary stress is hard to hear or pin down by phonometric methods.

(7) "The variation in my data […] is the result of the speech produced by three native speakers. […] All speakers showed variation, i.e. at one stage a speaker produced a binary pattern, and at another stage a word with comparable morphological and syllabic structure was produced with a ternary pattern. This was true for all the speakers. In general, it was observed that of the three speakers, one speaker tended towards a binary pattern, one towards a ternary pattern and one had no obvious preference. But despite these tendencies, all speakers produced both binary and ternary patterns for words with comparable morphological and syllabic structure." (Elenbaas 1999:99)

(8) Why is this important? Secondary stress plays a crucial role in the rest of Finnish phonology and morphology:
  (a) allomorph selection (see e.g. Anttila 1997, Anttila and Cho 1998, Anttila and Revithiadou 2000)
  (b) segmental alternations (see e.g. Anttila 2002, Keyser and Kiparsky 1984, Kiparsky 2003, Ringen and Heinämäki 1999)

(9) In order to make progress, we need
  (a) A way of collecting stress intuitions from a large number of speakers.
  (b) A theory that can handle the variation and gradience evident in the data.

(10) Outline:
  (a) Study variation in Finnish secondary stress indirectly, based on its segmental consequences, using a large corpus of internet data.
  (b) Propose an analysis based on Optimality Theory (Prince and Smolensky 1993/2004) and the Multiple Grammars Theory of variation (Kiparsky 1993, Anttila 2007).
  (c) Show that the analysis makes quantitative predictions that are confirmed by the corpus data.
  (e) Show how the analysis explains away some apparent morpholexical conditions and how it extends to other segmental alternations.

(11) Implication for variation theory. There are two types of quantitative patterns:
  (a) Quantitative universals that are independent of rankings (= T-orders).
  (b) Quantitative particulars that depend on rankings.

2. Methodology

(12) The key observation: Secondary stress has systematic segmental consequences.

(13) *t* → ∅ between unstressed light syllables (Keyser and Kiparsky 1984, Kiparsky 2003). Variable *t*-deletion reflects variable stress.

(a)	/maa-i-ta/	mái.ta	'country-PL-PAR'
(b)	/talo-i-ta/	tá.lo.ja	'house-PL-PAR'
(c)	/professori-i-ta/	pró.fes.sò.re.ja ~ pró.fes.so.rèi.ta	'professor-PL-PAR'

(14) *t*-deletion is a useful stress diagnostic because it is reflected in spelling.

(15) We used QueryGoogle to collect systematic data on *t*-deletion patterns on Finnish web sites http://www.linguistics.ucla.edu/people/hayes/QueryGoogle/

(16) Step 1. All vowel-final nominal stems were extracted from an unabridged dictionary (Tuomi 1972), available at http://www.ling.helsinki.fi/uhlcs/.

(17) Result: 23,105 stems. 256 stems with problematic syllabification were excluded (e.g. *oikeusto*, *capriccio*). Result: 22,849 stems.

(18) Step 2. Each stem was annotated for the number of syllables.

SYLLABLES	FREQUENCY	SYLLABLES	FREQUENCY
1	57	6	415
2	5,952	7	80
3	8,801	8	10
4	5,768	9	3
5	1,762	10	1

(19) Step 3. Each stem was annotated for the weights and vowels of the last three syllables. [Stems with alternating syllable weights were marked and excluded.]

```
tunturisto S01 S:4 P:UHO
neulasisto S01 S:4 P:AHO
lihaksisto S01 S:4 P:HHO
koordinaatisto S01 S:5 P:IHHO
konsonantisto S01 S:5 P:OHHO
luisto S01 S:2 P:HO
```

(20) Step 4. All hypothetically possible partitive plural forms for all stems were generated. This involves playing GEN (cf. Karttunen 2006). For example, consider the candidate forms for *Esperanto* 'Esperanto' and *salaatti* 'salad':

```
esperantoita S01 S:4 P:EHO V:T     (AA: really bad)
esperannoita S01 S:4 P:EHO V:T     (AA: really bad)
esperantoja S01 S:4 P:EHO V:0      (AA: good)

salaatteita S06 * S:3 P:HI V:T     (AA: bad)
salaateita S06 * S:3 P:HI V:T      (AA: sort of OK)
salaatteja S06 * S:3 P:HI V:0      (AA: good)
salaatteitä S06 * S:3 P:HI V:T     (AA: really bad)
salaateitä S06 * S:3 P:HI V:T      (AA: really bad)
salaattejä S06 * S:3 P:HI V:0      (AA: really bad)
```

(21) Step 5. The following stem types were excluded: (i) stems with alternating syllable weights; (ii) forms with questionable syllabification; (iii) stems ending in a heavy syllable (e.g. *lauantai*); (iv) forms with potentially problematic characters ({, |, $, %), including forms with vowels *ä, ö*. The last class is large, about 45% of the forms.

(22) Result:

- About 10,000 stems
- 29,547 potential partitive plural forms

(23) The QueryGoogle search was done on April 12, 2005, 10:44-11:12am.
Result: 9,283,597 actual partitive plural forms.

3. Empirical generalizations

(24) The key generalization: An extrametrical singleton *t* is deleted (cf. Anttila 2006)

(a)	/maa-i-ta/	(mái.ta)	'country-PL-PAR'
(b)	/talo-i-ta/	(tá.lo)ja	'house-PL-PAR'
(c)	/professori-i-ta/	(pró.fes.so)(rèi.ta) ~ (pró.fes)(sò.re)ja	'professor-PL-PAR'

3.1 Monosyllabic stems

(25) Monosyllabic stems show categorical *t*-retention.

(26)

(a)	*UNARY	No monosyllabic feet.
(b)	MAXφ	No deletion within a foot.
(c)	MAX	No deletion.
(d)	*T	No stops (**p, *t, *k*).

(27) Monosyllabic stems

/maa-i-ta/	*UNARY	MAXφ	MAX	*T
a. → (mái.ta)				*
b. (máa.ja)		*!	*	
c. (mái)ta	*!			*

3.2 Longer stems

(28) Generalization 1. The Weight-to-Stress Principle (WSP): Heavy syllables are stressed. This has two kinds of effects.

(29) Disyllabic stems: categorical *t*-deletion

/talo-i-ta/	...	WSP	MAX	*T
a. → (tá.lo)ja			*	
c. (tá.loi)ta		*!		*
d. (tá.loi.ta)		*!		*

(30) Longer stems: The skipping effect

(ád.re.na)(lìi.ne)ja	'adrenalin-PL-PAR'	initial dactyl
(ál.le)(gò.ri.o)ja	'allegory-PL-PAR'	initial trochee

/adrenaliini-i-ta/	…	WSP	MAX	*T
a. → (ád.re.na)(lìi.ne)ja			*	
b. (ád.re)(nà.lii.ne)ja		*!	*	
c. (ád.re)(nà.lii)(nèi.ta)		*!		*
d. (ád.re.na)(lìi.nei.ta)		*!		*
e. (ád.re.na)(lìi.nei)ta		*!		*
f. (ád.re)(nà.lii.nei)ta		*!*		*

(31) Generalization 2. CVV attracts stress from CVC (Karvonen 2005:81-94):
(kóor.di)(nàa.tis.to)ja CVV.CVC 'coordinate grid-PL-PAR'
(hó.ri.son)(tàa.le)ja CVC.CVV 'horizontal-PL-PAR'

(32) WSP/VV Heavy syllables with a long vowel are stressed.

/koordinaatisto-i-ta/	…	WSP/VV	WSP	MAX	*T
a.→ (kóor.di)(nàa.tis)(tòi.ta)			*		*
b.→ (kóor.di)(nàa.tis.to)ja			*	*	
c. (kóor.di.naa)(tìs.to)ja		*!	*	*	
/dialektiikka-i-ta/	…	WSP/VV	WSP	MAX	*T
a. (hó.ri)(sòn.taa)(lèi.ta)		*!	*		*
b. (hó.ri)(sòn.taa.le)ja		*!	*	*	
c. → (hó.ri.son)(tàa.le)ja			*	*	

(33) Stems longer than 2 syllables exhibit extensive variation. Different stems show very different quantitative profiles.

(34) Examples of variation. Three types of 4-syllable stems:
(a) (ál.ler.gi)(òi.ta) 99.7% ?(ál.ler)(gì.o)ja 0.3% 'allergy-PL-PAR'
(b) (pró.fes.so)(rèi.ta) 15.1% (pró.fes)(sò.re)ja 84.9% 'professor-PL-PAR'
(c) ?(fí.lo.so)(fèi.ta) 9.3% (fí.lo)(sò.fe)ja 90.7% 'philosopher-PL-PAR'

(35) Kiparsky (2003:113-4): Secondary stress is lexically fixed in some stems and variable in others. However, there remain subtle phonological conditions that govern the distribution of these fixed stresses.

(36) Generalization 3. The vowel sonority effect: /a, ä, o, ö/ are preferably stressed, /i, e, u, y/ are preferably unstressed (see also Kenstowicz 1994, de Lacy 2004, Crowhurst and Michael 2005). Morphophonemically {/a/, /o/} = /A/ and {/e/, /i/} = /I/.

(37) /allergia/ (X.H.I.A) 'allergy' vs. /professori/ (X.H.A.I) 'professor'
(a) ál.ler.gi.òi.ta 99.7% ál.ler.gì.o.ja 0.3% X.H.I.A.X
(b) pró.fes.so.rèi.ta 15.1% pró.fes.sò.re.ja 84.9% X.H.A.I.X

(38) Generalization 4. The prominence clash effect: Secondary stress avoids falling next to a heavy syllable (Inkelas 1999, 145).

(39) /professori/ (X.H.A.I) 'professor' vs. /filosofi/ (X.L.A.I) 'philosopher'
(a) pró.fes.so.rèi.ta 15.1% pró.fes.sò.re.ja 84.9% X.H.A.I.X
(b) fí.lo.so.fèi.ta 9.3% fí.lo.sò.fe.ja 90.7% X.L.A.I.X

(40) The constraints:

(a)	*x/A	No unstressed short A (= /a, ä, o, ö/)
(b)	*x/AA	No unstressed long A (= /a, ä, o, ö/)
(c)	*X/I	No stressed short I (= /e, i, u, y/)
(d)	*X/II	No stressed long I (= /e, i, u, y/)
(e)	*H.X	No stress next to a heavy syllable.

3.3 A morphological generalization

(41) Generalization: *t*-deletion only applies to suffixes, never to stems.

(42) PARSE-STEM Stems are exhaustively footed (undominated).
/traditio-i-ta/ → (trá.di.**t**i)(òi.ta) 'tradition-PL-PAR'

3.4 Other generalizations

(43) The following constraints are familiar from earlier work:

(a)	SWP	Stressed syllables are heavy.
(b)	FTBIN	Feet are disyllabic.
(c)	PARSE-σ	Syllables belong to feet.
(d)	ALL-FEET-LEFT	All feet are at the left edge of the prosodic word.

4. Analysis

(44) (a) Work out the predictions made by the above constraints.
(b) See how well they approximate the observed patterns in the corpus.

4.1 Three-syllable stems

(45) Possible 3-syllable stems (4 types): X.{L,H}.{I,A}

TYPE	GHITS	EXAMPLE	GLOSS
HI	1,804,407	po.lii.si	'police'
LI	1,180,225	ka.me.li	'camel'
HA	2,205,722	kor.jaa.mo	'repair shop'
LA	2,030,424	ka.me.ra	'camera'

(46) Constraint violation profiles for 3-syllable stems. Candidates that violate the constraints *UNARY, MAXφ and PARSE-STEM have been suppressed. The constraints WSP/VV and WSP are undominated.

		WSP/VV	WSP	FTBIN	*T	SWP	*X.H	*x/A	*x/AA	*X/I	*X/II	ALIGN-L	P-SYLL	MAX
X.H.I.X	(po.lii)(sei.ta)	1	1		1	1	2	1			1	2		
	(po.lii.sei)ta	2	2	1	1	1	1	1					1	
	(po.lii.se)ja	1	1	1		1	1	1					1	1
X.L.I.X	(ka.me)(lei.ta)				1	1		1			1	2		
	(ka.me.lei)ta	1	1	1	1	1		1					1	
	(ka.me.le)ja			1		1		1					1	1
X.H.A.X	(kor.jaa)(moi.ta)	1	1		1		2	1	1			2		
	(kor.jaa.moi)ta	2	2	1	1		1	1	2				1	
	(kor.jaa.mo)ja	1	1	1			1	2	1				1	1
X.L.A.X	(ka.me)(roi.ta)				1	1		1				2		
	(ka.me.roi)ta	1	1	1	1	1		1	1				1	
	(ka.me.ro)ja			1		1		2					1	1

(47) Two foot structures are predicted to be possible:

(a)	(′x.x)(x.x)	(po.lii)(sei.ta)	no t-deletion
(b)	(′x.x.x)x	(po.lii.se)ja	t-deletion

(48) Factorial typology computed by OTSOFT (Hayes, Tesar and Zuraw 2003). Winners with *t*-deletion are grayed out.

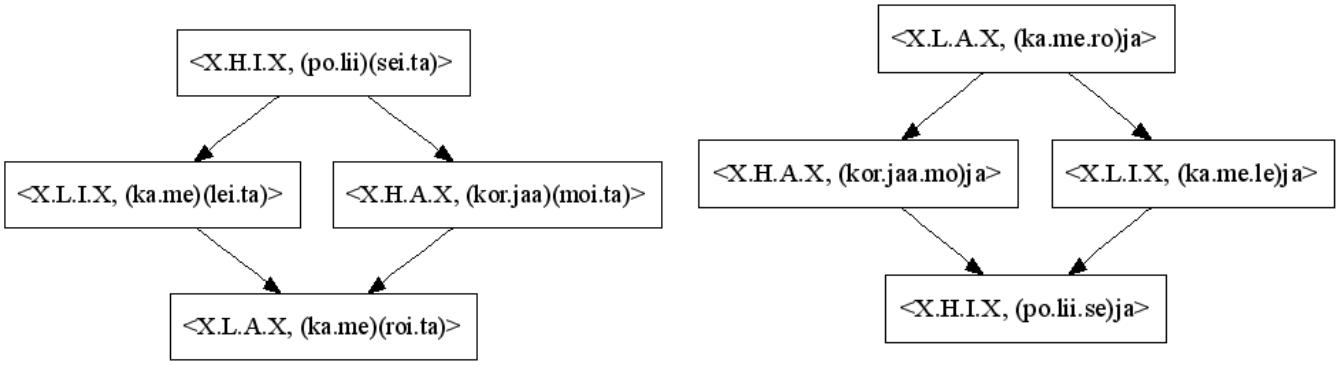
Output #1	Output #2	Output #3	Output #4
(po.lii)(sei.ta)	(po.lii.se)ja	(po.lii.se)ja	(po.lii.se)ja
(ka.me)(lei.ta)	(ka.me)(lei.ta)	(ka.me)(lei.ta)	(ka.me.le)ja
(kor.jaa)(moi.ta)	(kor.jaa)(moi.ta)	(kor.jaa.mo)ja	(kor.jaa)(moi.ta)
(ka.me)(roi.ta)	(ka.me)(roi.ta)	(ka.me)(roi.ta)	(ka.me)(roi.ta)

Output #5	Output #6
(po.lii.se)ja	(po.lii.se)ja
(ka.me.le)ja	(ka.me.le)ja
(kor.jaa.mo)ja	(kor.jaa.mo)ja
(ka.me)(roi.ta)	(ka.me.ro)ja

(49) Implicational universals.
- (a) *t*-deletion in /kamera/ (LA) implies deletion everywhere.
- (b) *t*-deletion in /kameli/ (LI) implies deletion in /poliisi/ (HI).
- (c) etc.

(50) We used T-ORDER GENERATOR (Anttila and Andrus 2006) to compute all the implicational universals hidden in the factorial typology.

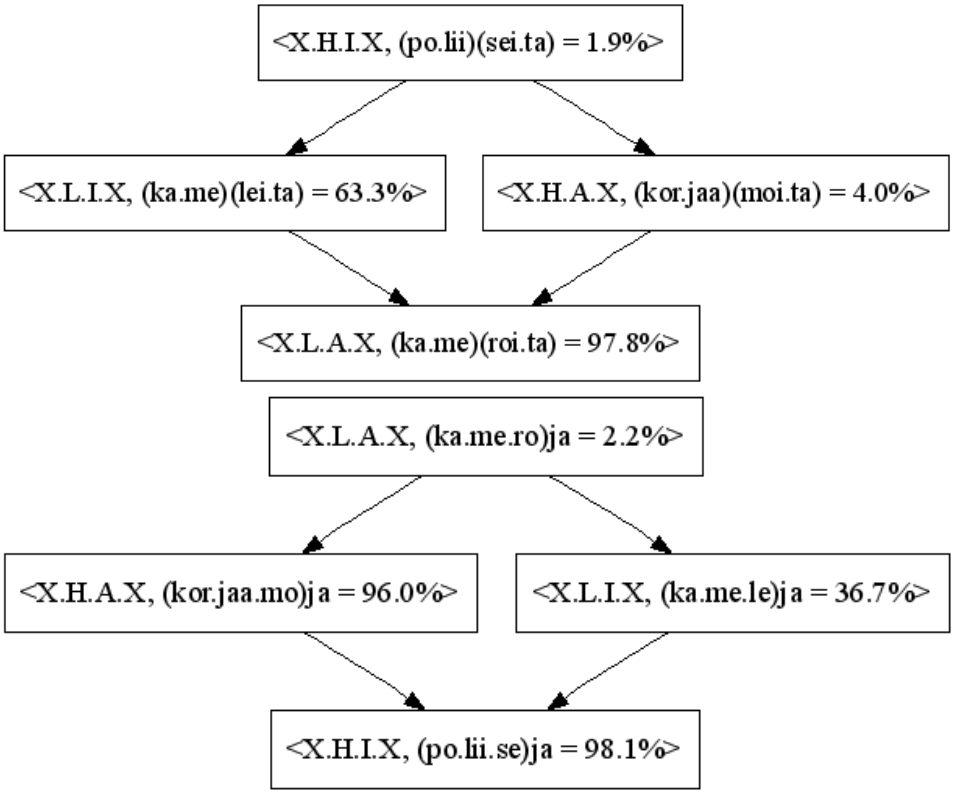
(51) T-order for trisyllabic stems



(52) Recall the Multiple Grammars Theory of variation (Kiparsky 1993, Anttila 2007):
(a) Variation arises from multiple grammars within/across individuals.
(b) The number of grammars predicting an output is proportional to its frequency of occurrence.

(53) A quantitative prediction: The observed frequencies should grow in the direction of the arrows.

(54) The T-order annotated with observed frequencies **[SLIDE]**



(55) Recall that these predictions follow under Multiple Grammars (e.g. Kiparsky 1993), Partially Ordered Grammars (e.g. Anttila and Cho 1998) and Stochastic OT (e.g. Boersma and Hayes 2001) because the factorial typology is the same.

(56) How well are we doing?
(a) Precision = 1.0
(b) Recall = 0.417

(57) A note on postlexical stress. The analysis only predicts a subset of the secondary stresses discussed in the literature. For example, the optional final stress in *rá.vin.tò.lat ~ rá.vin.to.làt* 'restaurants' is not predicted.

(58) Proposal: In addition to word-level stress, there are postlexical stresses (and other prominences) that do not interact with segmental phonology.

4.2 Four-syllable stems

(59) Possible 4-syllable stems (12 types): X.{L,H}.{I,A,H}.{I,A}

TYPE	GHITS	EXAMPLE	GLOSS
HAI	34,612	pro.fes.so.ri	'professor'
LAI	23,595	fi.lo.so.fi	'philosopher'
HII	630,594	ar.tik.ke.li	'article'
LII	4,175	py.ra.mi.di	'pyramid'
HAA	24,479	ra.vin.to.la	'restaurant'
LAA	30	ta.pi.o.la	'Tapiola'
HIA	190,416	al.ler.gi.a	'allergy'
LIA	91,598	gal.le.ri.a	'gallery'
HHA	92,308	e.dus.tus.to	'representation'
HHI	80,063	ter.mos.taat.ti	'thermostat'
LHA	13,039	af.fri.kaat.ta	'affricate'
LHI	392,942	mar.ga.rii.ni	'margarine'

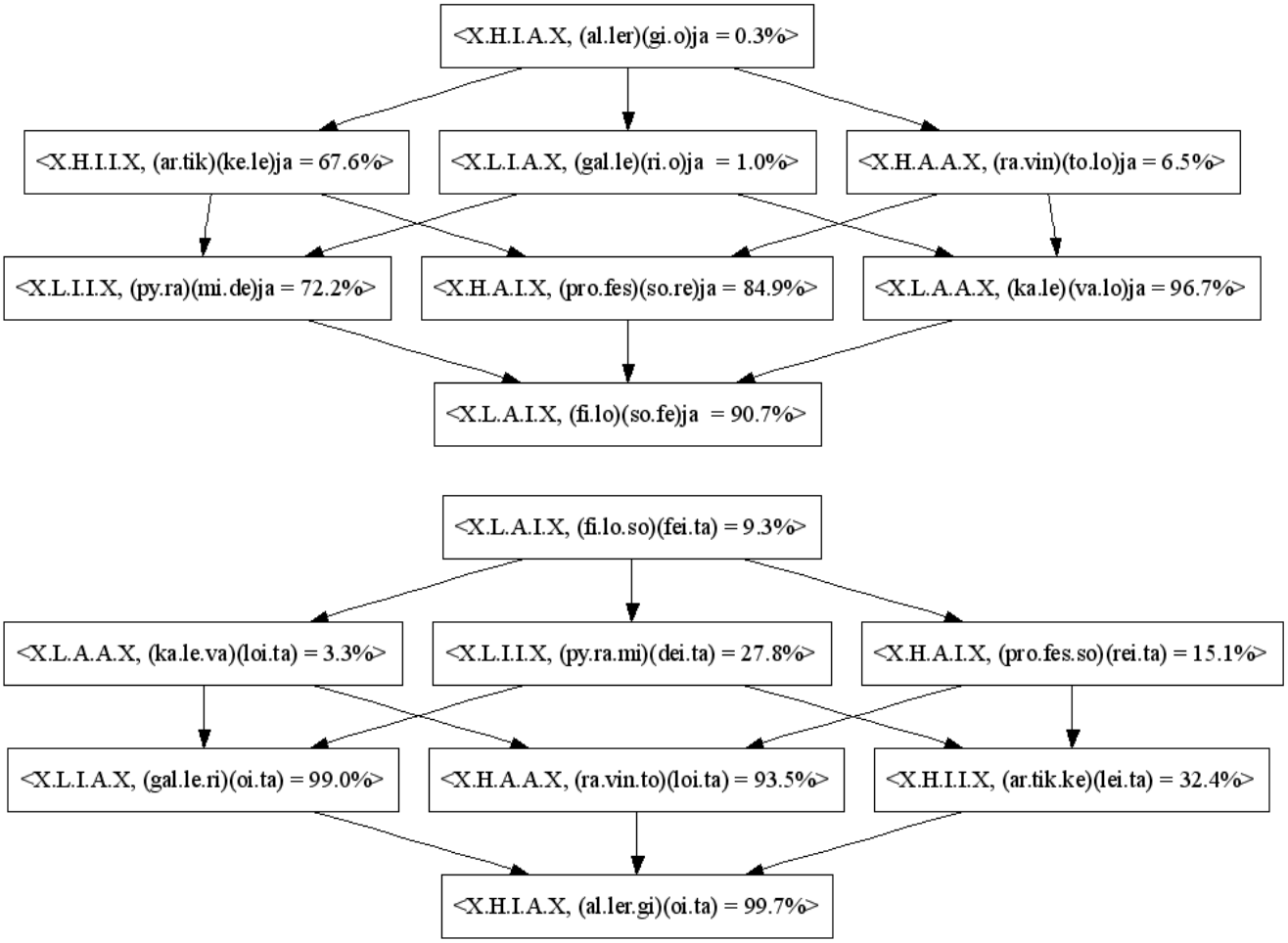
(60) Two foot structures are predicted to be possible:

(a)	(′x.x)(x.x)x	(pro.fes)(so.re)ja	t-deletion
(b)	(′x.x.x)(x.x)	(pro.fes.so)(rei.ta)	no t-deletion

(61) *t*-deletion is predicted to be categorical for the following input types:

TYPE	DELETION OUTPUT	DELETION%	NON-DELETION TOKENS
HHA	(e.dus)(tus.to)ja	100.0%	22
HHI	(ter.mos)(taat.te)ja	100.0%	23
LHA	(aff.ri)(kaat.to)ja	99.7%	43
LHI	(mar.ga)(rii.ne)ja	100.0%	138

(62) T-order for the remaining stems. Precision = 0.956, recall = 0.586 **[SLIDE]**



(63) Evaluation: The binary pattern (*ka.le*)(*va.lo*)*ja* (X.L.A.A.X) occurs at a rate higher than predicted. There are only 7 such stems (30 tokens): *autostrada*, *eldorado, subimago, tapiola, telefoto, tuatara, unisono.*

(64) Diagnosis: Some of these words are compounds, e.g. *auto=strada*, *tele=foto, uni=sono*, in which case only binarity is predicted.

4.3 Five-syllable stems

(65) Possible 5-syllable stems (36 types, 28 exist): X.{L,H}.{I,A,H}.{I,A,H}.{I,A}.X.

TYPE	GHITS	EXAMPLE	GLOSS
HHIA	10,126	konsultaatio	‘consultation’
LAIA	63,717	allegoria	‘allegory’
LHIA	42,237	operaatio	‘operation’
LIIA	3,474	kommunikea	‘communiqué’
HAIA	21	responsorio	‘responsory’
HIIA	221,456	opiskelija	‘student’
LHAI	20,316	operaattori	‘operator’
LHII	2,139	inkunaabeli	‘incunable’
HHAI	1,525	kommentaattori	‘commentator’
HHII	43	liirumlaarumi	‘nonsense’
LAII	69	polyamidi	‘polyamide’
LAAI	1,990	antropologi	‘anthropologist’
LIII	296	symposiumi	‘symposium’
HIII	6	kompendiumi	‘compendium’
LIAI	66	jeremiadi	‘jeremiad’
LHHA	207	koordinaatisto	‘coordinate grid’
HAAI	23	konkvistadori	‘conquistador’
LHHI	1,128	akvarellisti	‘aquarellist’
HHHI	153	avantgardisti	‘avant-gardist’
HIAI	50	kolesteroli	‘cholesterol’
HIHA	290	evankelista	‘evangelist’
LIHI	103,312	auktoriteetti	‘authority’
HAHA	14	edustajisto	‘representation’
LAHI	9,225	adrenaliini	‘adrenalin’
LAHA	167	matematiikka	‘mathematics’
HIHI	1,884	republikaani	‘republican’
HAHI	969	konservatiivi	‘conservative’
LIHA	65	italiaano	‘Italian (person)’

(66) I have found no stems of the following 8 types in the dictionary (Tuomi 1972):

LAAA	HAAA	HAII	LIAA
HIAA	HHHA	LHAA	HHAA

(67) Three foot structures are predicted to be possible:

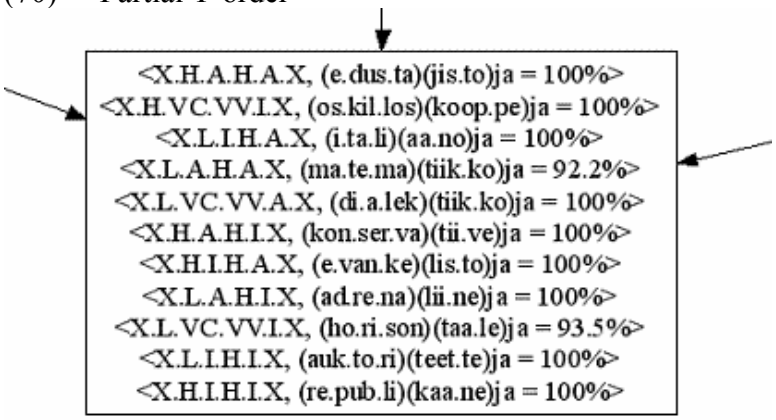
(a)	(′x.x.x)(x.x)x	(ad.re.na)(lii.ne)ja	t-deletion
(b)	(′x.x)(x.x)(x.x)	(al.le)(go.ri)(oi.ta)	no t-deletion
(c)	(′x.x)(x.x.x)x	(ak.va)(rel.lis.te)ja	t-deletion

(68) T-order: precision = 0.857, recall = 0.633 **[SLIDE]**

(69) Type 1: Initial dactyl, categorical *t*-deletion (11 types).

(ad.re.na)(lii.ne)ja	LAHI	
(ma.te.ma)(tiik.ko)ja	LAHA	
(i.ta.li)(aa.no)ja	LIHA	
(auk.to.ri)(teet.te)ja	LIHI	
(di.a.lek)(tiik.ko)ja	LHHA	(CVC.CVV)
(ho.ri.son)(taa.le)ja	LHHI	(CVC.CVV)
(e.dus.ta)(jis.to)ja	HAHA	
(kon.ser.va)(tii.ve)ja	HAHI	
(e.van.ke)(lis.to)ja	HIHA	
(re.pub.li)(kaa.ne)ja	HIHI	
(os.kil.los)(koop.pe)ja	HHHI	(CVC.CVV)

(70) Partial T-order



(71) Evaluation: In these stems, no *t*-deletion is predicted, but some is observed:

LAHA	automatiikoita (11), matematiikoita (1), pedagogiikoita (1)
LHHI	arkkitehtuureita (65), instrumentaaleita (3)
LAHI	alkoholisteita (1), automobiileita (3)
LIHI	materiaaleita (4), semifinaaleita (14), vokalisoinneita (1)

(72) In a number of cases, the predicted initial dactyl appears intuitively incorrect, e.g.

LAHI	(he.mo)(glo.bii.ne)ja	'haemoglobin-PL-PAR'	
	(her.ma)(fro.diit.te)ja	'hermaphrodite-PL-PAR'	
	(lak.to)(ba.sil.le)ja	'lactobacillus-PL-PAR'	
LIHI	(mo.no)(fy.siit.te)ja	'monophysite-PL-PAR'	(2 hits)
	cf. (mo.no.te)(is.te)ja	'monotheist-PL-PAR'	(21 hits)
	(an.ti)(se.miit.te)ja	'anti-semite-PL-PAR'	(2 hits)
	cf. (an.ti.bi)(oot.te.ja)	'antibiotic-PL-PAR'	(861 hits)

(73) Possible diagnoses:

(a) Compounds: *árkki=tèhtuuri*, *álko=hòlisti*, *áuto=mòbiili*, *sémi=fînaali*.

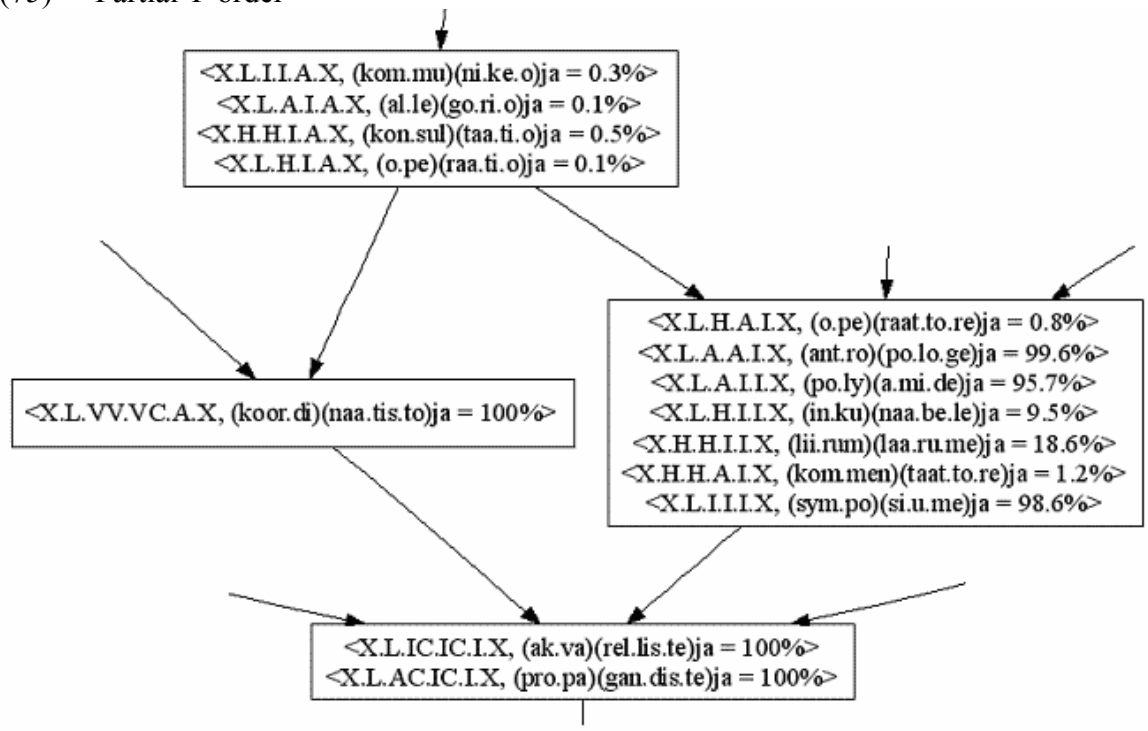
(b) Some disyllabic suffixes (e.g. *-ìsti*) may be independent prosodic words

(c) Lexical frequency: compound stress seems common in infrequent words.

(74) Type 2: Initial trochee, variable *t*-deletion (14 types).

(ant.ro)(po.lo)(gei.ta)	~	(ant.ro)(po.lo.ge)ja	LAAI	
(po.ly)(a.mi)(dei.ta)	~	(po.ly)(a.mi.de)ja	LAII	
(al.le)(go.ri)(oi.ta)	~	(al.le)(go.ri.o)ja	LAI	
(kom.mu)(ni.ke)(oi.ta)	~	(kom.mu)(ni.ke.o)ja	LIIA	
(sym.po)(si.u)(mei.ta)	~	(sym.po)(si.u.me)ja	LIII	
(o.pe)(raat.to)(rei.ta)	~	(o.pe)(raat.to.re)ja	LHAI	
(o.pe)(raa.ti)(oi.ta)	~	(o.pe)(raa.ti.o)ja	LHIA	
(in.ku)(naa.be)(lei.ta)	~	(in.ku)(naa.be.le)ja	LHII	
(koor.di)(naa.tis)(toi.ta)	~	(koor.di)(naa.tis.to)ja	LHHA	(CVV.CVC)
(pro.pa)(gan.dis)(tei.ta)	~	(pro.pa)(gan.dis.te)ja	LHHI	(CAC.CIC)
(ak.va)(rel.lis)(tei.ta)	~	(ak.va)(rel.lis.te)ja	LHHI	(CIC.CIC)
(kom.men)(taat.to)(rei.ta)	~	(kom.men)(taat.to.re)ja	HHAI	
(lii.rum)(laa.ru)(mei.ta)	~	(lii.rum)(laa.ru.me)ja	HHII	
(kon.sul)(taa.ti)(oi.ta)	~	(kon.sul)(taa.ti.o)ja	HHIA	

(75) Partial T-order



(76) Evaluation: The box on the right has both very high and very low values. Some intuitively incorrect stress patterns are predicted as well, e.g. *(*ant.ro*)(*po.lo.ge*)*ja* instead of (*ant.ro.po*)(*lo.ge*)*ja*

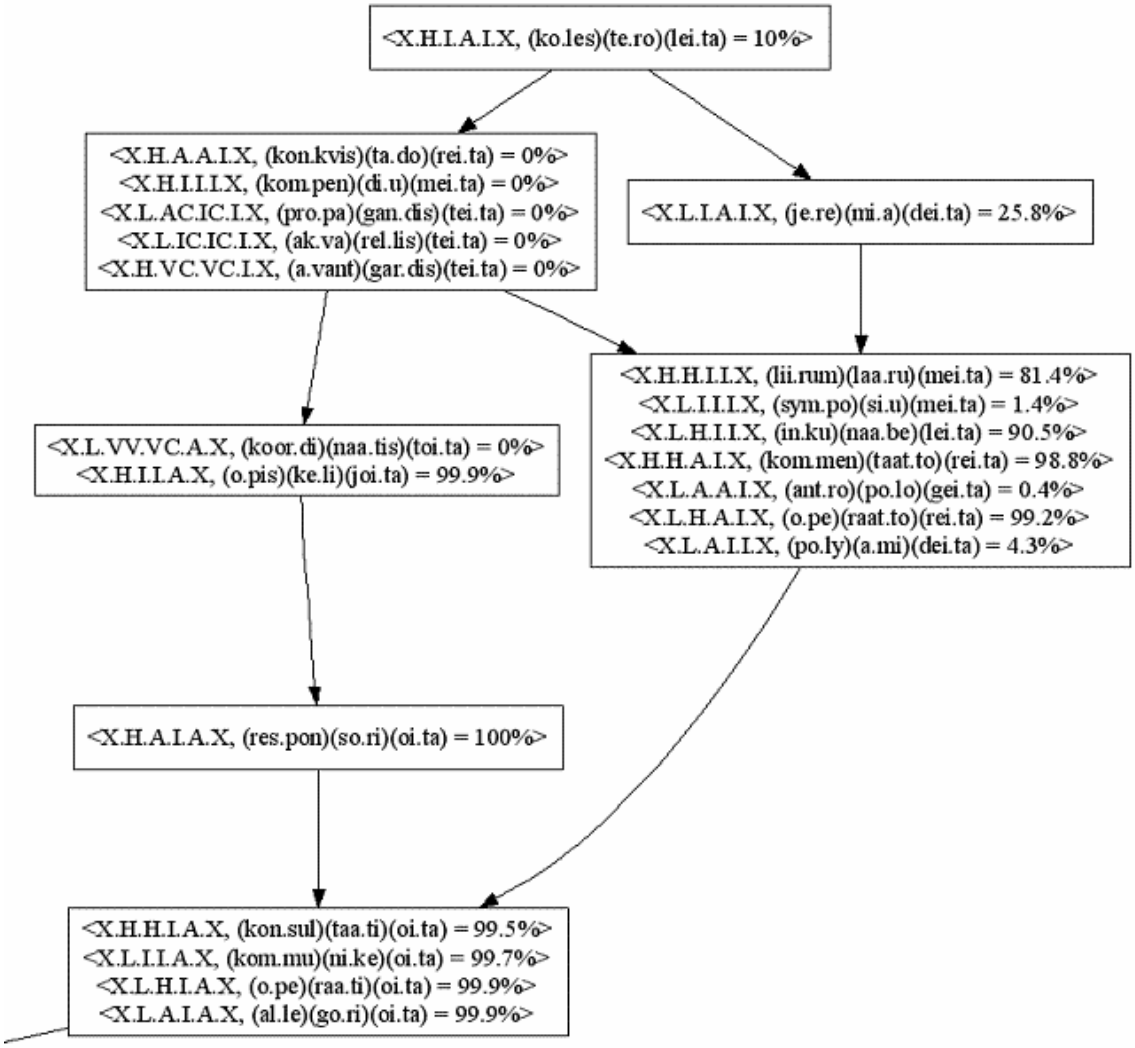
(77) Diagnosis: Some learned disyllabic prefixes and suffixes, e.g. *mono-* 'mono-', *poly-* 'poly-', *-logi* '-logist', *-grafi* '-grapher', may be independent prosodic words. Again, we seem to have compounding.

(78) Prediction: /-lògi/ should avoid occurring next to a stressed syllable. This seems correct. Four apparent counterexamples: *oologi* 'oologist', *zoologi* 'zoologist' (quadrisyllables?), *prologi* 'prologue' and *eklogi* 'eclogue' (different suffix).

(79) Type 3: Initial trochee or dactyl, variable *t*-deletion (7 types)

(je.re)(mi.a)(dei.ta)	~ (je.re.mi)(a.de)ja	~ (je.re)(mi.a.de)ja	LIAI
(res.pon)(so.ri)(oi.ta)	~ (res.pon.so)(ri.o)ja	~ (res.pon)(so.ri.o)ja	HAIA
(kon.kvis)(ta.do)(rei.ta)	~ (kon.kvis.ta)(do.re)ja	~ (kon.kvis)(ta.do.re)ja	HAAI
(ko.les)(te.ro)(lei.ta)	~ (ko.les.te)(ro.le)ja	~ (ko.les)(te.ro.le)ja	HIAI
(o.pis)(ke.li)(joi.ta)	~ (o.pis.ke)(li.jo)ja	~ (o.pis)(ke.li.jo)ja	HIIA
(kom.pen)(di.u)(mei.ta)	~ (kom.pen.di)(u.me)ja	~ (kom.pen)(di.u.me)ja	HIII
(a.vant)(gar.dis)(tei.ta)	~ (a.vant.gar)(dis.te)ja	~ (a.vant)(gar.dis.te)ja	HHHI

(80) Partial T-order



(81) Evaluation: Some phonologically similar stems differ in terms of initial trochee vs. initial dactyl:

(a)	(in.ter)(fe.ro)(nei.ta) ~ *(in.ter.fe)(ro.ne)ja	~ (in.ter)(fe.ro.ne)ja	HIAI
	(tes.tos)(te.ro)(nei.ta) ~ *(tes.tos.te)(ro.ne)ja	~ (tes.tos)(te.ro.ne)ja	HIAI
(b)	*(ko.les)(te.ro)(lei.ta) ~ (ko.les.te)(ro.le)ja	~ *(ko.les)(te.ro.le)ja	HIAI

(82) Diagnosis: (*inter*-) and (*testos*-) appear to be independent prosodic words, whereas (*koles*) is not.

(83) In the last class of 5-syllables, footing cannot be unambiguously inferred from the segmental structure: (*je.re.mi*)(*a.de*)*ja* and (*je.re*)(*mi.a.de*)*ja* 'jeremiad-PL-PAR' are both predicted. These forms are not included in precision and recall figures.

5. Eliminating "preaccentuation"

(84) Kiparsky (2003): Some stems have fixed stress (underlined), others mobile stress:

(a)	Alabama	'Alabama'	(b)	Amerikka	'America'
	barrikadi	'barricade'		kalenteri	'calendar'

(85)

(a) No skipping	(b) Optional skipping
(á.la)(bà.mo)ja / *(á.la)ba(mòi.ta)	(á.me)(rìk.ko)ja ~ (á.me)ri(kòi.ta)
(bár.ri)(kà.de)ja / *(bár.ri)ka(dèi.ta)	(ká.len)(tè.re)ja ~ (ká.len)te(rèi.ta)

(86) Response: The difference is phonological (vowel sonority, prominence clash).

(87) Certain light inflectional endings have traditionally been considered preaccenting (Sadeniemi 1949, Carlson 1978, Hanson and Kiparsky 1996, Elenbaas 1999):

(a)	ó.pet.ta.jà.na	'teacher-ESS'	/-nA/	'essive'
(b)	ó.pet.ta.jà.ni	'teacher-1.SG.POSS'	/-ni/	'1. sg. possessive'
(c)	jý.räh.te.lè.vä	'boom-PRES.PCP'	/-vA/	'present participle'

(88) Response: The generalization seems flawed. In many of the examples, the second syllable is heavy and/or the fourth syllable has a low vowel:

(a)	ó.**pet**.ta.jà.na	??ó.**pet.**tà.ja.na	H.A.A	'teacher-ESS'
	??ká.**le**.va.là.na	ká.**le**.và.la.na	L.A.A	'Kalevala-ESS'
(b)	kómp.pa.ni.**à**.na	??kómp.pa.nì.**a**.na	L.I.A	'company-ESS'
	??fí.lo.so.**fi**.na	fí.lo.sò.**fi**.na	L.A.I	'philosopher-ESS'

(89) Data from Elenbaas (1999, Appendix B, pp. 207-216):

- 294 Finnish words, impressionistically stress-marked by Elenbaas
- 166 "preaccenting" suffixes, i.e. CV suffixes that do not close the preceding syllable, but appear to cause the preceding syllable to be stressed

(90) Preview of results: The case for preaccentuation rests on 11 words, all with optional preaccentuation. In all other cases, there is an alternative metrical explanation available.

(91) 1st and 2nd person singular possessive suffixes /-ni, -si/, 88 examples

(a) No stress before the possessive suffix (32)

mér.ko.nò.mi.ni	'business.degree.holder-POSS.1.SG'
pú.he.li.mès.ta.ni	'telephone-ELA-POSS.1.SG'

(b) Stress in the rhythmically expected position (40)

á.vai.mèe.ni	'key-ILL-POSS.1.SG'
má.te.ma.tìi.kas.sà.ni	'mathematics-ESS-POSS.1.SG'

(b) Prominence clash effect (5)

á.**vai**.mi.nà.ni	'key-ELA-POSS.1.SG'
kún.**nal**.li.sè.ni	'communal-POSS.1.SG'

(d) Sonority effect (5)

á.te.**ri.à**.ni	'meal-POSS.1.SG'
váa.ti.màt.to.**mi.nà**.ni	'humble-PL-ESS-POSS.1.SG'

(e) Potential examples of preaccentuation, all optional (6)

pú.he.li.mè.ni	'telephone-POSS.1.SG'
pé.ri.jä.nå.ni	'heir-ESS-POSS.1.SG'
ér.go.nò.mi.a.nà.ni	'ergonomy-ESS-POSS.1.SG'
ó.pis.kè.li.ja.nà.ni	'student-ESS-POSS.1.SG'
pá.ki.nòit.si.ja.nà.ni	'columnist-ESS-POSS.1.SG'
váa.ti.màt.to.ma.nà.ni	'humble-ESS-POSS.1.SG'

(92) Essive case /-na, -nä/, 71 examples

(a) No stress before the essive (24)

mér.ko.nò.mi.na	'business.degree.holder-ESS'
má.te.ma.tìik.ka.nà.ni	'mathematics-ESS-POSS.1.SG'

(c) Stress in the rhythmically expected position (27)

káu.pun.kì.na	'city-ESS'
ér.go.nò.mi.à.na	'ergonomy-ESS'
rá.kas.ta.jàt.ta.rè.na.ni	'lover-FEM-ESS-POSS.1.SG'

(c) Prominence clash effect (13)

kún.**nal**.li.sè.na	'communal-ESS'
tár.**jot**.ti.mè.na	'tray-ESS'

(d) Sonority effect (3)

á.te.**ri.à**.na.ni	'meal-ESS-POSS.1.SG'

(e) Potential examples of preaccentuation, all optional (4)

pú.he.li.mè.na	'telephone-ESS'
pú.he.li.mì.na	'telephone-PL-ESS'
pú.he.li.mè.na.ni	'telephone-ESS-POSS.1.SG'
pú.he.li.mè.nam.me	'telephone-ESS-POSS.1.PL'

(93) The partitive case /-(t)a, (t)ä/, 51 examples

(a) No stress before the partitive (25)

kún.nal.lì.si.a	‘communal-PL-PAR’
pú.he.lì.mi.à.ni	‘telephone-PL-PAR-POSS.1.SG’

(b) Stress in the rhythmically expected position (25)

váa.ti.màt.to.mì.a.ni	‘humble-PL-PAR-POSS.1.SG’
pái.men.to.làis.ta.ni	‘nomad-PAR-POSS.1.SG’

(c) Potential examples of preaccentuation, optional (1)

káu.pun.kì.am.me (?)	‘city-PAR-POSS.1.PL’

(94) Conclusion: The case for preaccentuation is weak.

6. Different solutions to the same metrical problem

(95) The same metrical problem is resolved differently in different morphological constructions.

(96) Deletion: Infinitive /-taC/: *d* ~ ∅

(a)	/juo-taC/	(*juo.da*)	‘drink-INF’
(b)	/puno-taC/	(*pu.no*)*a*	‘weave-INF’
(c)	/vakioi-taC/	(*va.ki*)(*oi.da*)	‘keep.constant-INF’
(d)	/formalisoi-taC/	(*for.ma*)*li*(*soi.da*)	‘formalize-INF’

(97) Assibilation (Anttila 2003, 2006), Past Tense: *t* ~ *s*

(a)	/vetä-i/	(*ve.ti*)	‘pull-PAST’
(b)	/huuta-i/	(*huu.ti*)~(*huu*)*si*	‘shout-PAST’
(c)	/tilat-i/	(*ti.la*)*si*	‘order-PAST’

(98) Gemination: Past Participle /-tu/: *t* ~ *tt*

(a)	/juo-tu/	(*juo.tu*)	‘drink-PAST.PCP’
(b)	/puno-tu/	(*pu.not*)*tu*	‘weave-PAST.PCP’
(c)	/vakioi-tu/	(*va.ki*)(*oi.tu*)	‘keep.constant-PAST.PCP’
(d)	/formalisoi-tu/	(*for.ma*)*li*(*soi.tu*)	‘formalize-PAST.PCP’

(99) Gemination: Passive /-ta/: *t* ~ *tt*

(a)	/juo-ta-va/	(*juo.ta*)*va*	‘drink-PASS-PCP’
(b)	/puno-ta-va/	(*pu.not*)*ta.va*	‘weave-PASS-PCP’
(c)	/vakioi-ta-va/	(*va.ki*)(*oi.ta*)*va*	‘keep.constant-PASS-PCP’
(d)	/formalisoi-ta-va/	(*for.ma*)*li*(*soi.ta*)*va*	‘formalize-PASS-PCP’

(100) Gemination is also found in stem nativization. Stems must end in CV (Keyser and Kiparsky 1984), hence *i*-epenthesis. In addition, the preceding /p, t, k/ geminates:

Russell → (*Rus.se*)*li*	*gallup* → (*gal.lup*)*pi*	*(*gal.lu*)*pi*
Bresnan → (*Bres.na*)*ni*	*Traugott* → (*Trau.got*)*ti*	*(*Trau.go*)*ti*
Pullum → (*Pul.lu*)*mi*	*Ostyak* → (*ost.jak*)*ki*	*(*ost.ja*)*ki*

(101) Phonotactic gap: All noun stems

(a)	(*le.o*)	(*mu.ta*)	(*mut.ta*)	'Leo' / 'mud' / 'but'
	(*di.a*)	(*ku.ka*)	(*kuk.ka*)	'slide' / 'who' / 'flower'
(b)	(*lapi*)*o*	*(*si.ne*)*ti*	(*sinet*)*ti*	'spade'/ -- / 'seal'
	(*hope*)*a*	*(*lompa*)*ko*	(*lompak*)*ko*	'silver' / -- / 'wallet'

(102) Summary: The same metrical problem is repaired differently in different morphological constructions suggesting that they have different rankings.
- (a) Lenition: partitive (case), infinitive, past tense
- (b) Fortition: past participle, passive, nativization
- (c) Phonotactic gap: all noun stems

(103) In the case of stop deletion, the low-ranking faithfulness constraint was MAX. In other morphological contexts, it appears to be some other faithfulness constraint.

(104) Conjecture (Kiparsky?): The ranking of markedness constraints is the same across morphological constructions; the ranking of faithfulness constraint varies.

(105)
- (a) DEP(μ) >> MAX(T) deletion
- (b) MAX(T) >> DEP(μ) gemination
- (c) etc.

(106) The hard question: What determines the choice?

7. Summary

(107) Optimality Theory predicts typological generalizations (T-ORDERS)

(108) These generalizations are reflected quantitatively in language variation

(109) Evidence: Variable word stress in Finnish. The relevant factors include at least:
- syllable weight (CVV/CVC vs. CV)
- syllable closure (CVV vs. CVC)
- vowel sonority (/a, o/ vs. /i, e/)
- prominence clash (weight, stress)
- morphology (compounding, lexically accented suffixes)
- lexical frequency?

Appendix

Stress patterns of 5-syllable stems: initial dactyl, initial trochee, variation. Only forms that exhibit *t*-deletion are included.

`elaine:~/finstress/five-syllable-differences.txt`.

References

Anttila, Arto. 1997. 'Deriving variation from grammar', in Frans Hinskens, Roeland van Hout, and Leo Wetzels (eds.), *Variation, Change and Phonological Theory*, John Benjamins, Amsterdam / Philadelphia, pp. 35-68.

Anttila, Arto. 2002. 'Morphologically Conditioned Phonological Alternations', *Natural Language and Linguistic Theory* 20, 1-42. Also on http://roa.rutgers.edu/.

Anttila, Arto. 2006. 'Variation and opacity', *Natural Language and Linguistic Theory* 24, 893-944. Also on http://roa.rutgers.edu/.

Anttila, Arto. 2007. 'Variation and optionality', in Paul de Lacy, (ed.), *The Cambridge Handbook of Phonology*, Cambridge University Press, Cambridge, pp. XXX-XXX

Anttila, Arto and Curtis Andrus. 2006. 'T-orders' (comes with software), available at http://www.stanford.edu/~anttila/research/torders/t-order-manual.pdf.

Anttila, Arto and Young-mee Yu Cho. 1998. 'Variation and change in Optimality Theory', *Lingua* 104, 31-56, Special issue on Conflicting Constraints.

Boersma, Paul, and Bruce Hayes. 2001. 'Empirical tests of the Gradual Learning Algorithm', *Linguistic Inquiry* 32, 45-86.

Carlson, Lauri. 1978. 'Word stress in Finnish', Ms., Massachusetts Institute of Technology, Cambridge, Massachusetts.

Crowhurst, Megan J. and Lev Michael. 2005. 'Iterative footing and prominence driven stress in Nanti (Kampa)', *Language* 81, 47-95.

de Lacy, Paul. 2004. 'Markedness conflation in Optimality Theory', *Phonology* 21, 145-199.

Elenbaas, Nine. 1999. *A Unified Account of Binary and Ternary Stress*. Doctoral dissertation, University of Utrecht.

Elenbaas, Nine, and René Kager. 1999. 'Ternary Rhythm and the Lapse Constraint', *Phonology* 16, 273-329.

Hanson, Kristin, and Paul Kiparsky. 1996. 'A parametric theory of poetic meter', *Language* 72, 287-335.

Hayes, Bruce, Bruce Tesar, and Kie Zuraw. 2003. 'OTSoft 2.1, software package', available at http://www.linguistics.ucla.edu/people/hayes/otsoft/.

Inkelas, Sharon. 1999. 'Exceptional stress-attracting suffixes in Turkish: representations vs. the grammar', in René Kager, Harry van der Hulst & Wim Zonneveld, (eds.), *The Prosody-Morphology Interface*, Cambridge University Press, Cambridge, pp.134-87.

Karttunen, Lauri. 2006. 'The Insufficiency of Paper-and-Pencil Linguistics: the Case of Finnish Prosody'. ROA-818.

Karvonen, Dan. 2005. *Word Prosody in Finnish*. Ph.D. thesis, University of California at Santa Cruz.

Kenstowicz, Michael. 1994. 'Sonority-driven stress', ROA-33.

Keyser, Samuel Jay and Paul Kiparsky. 1984. 'Syllable structure in Finnish Phonology', in Mark Aronoff and Richard Oehrle, (eds.), *Language Sound Structure*, MIT Press, Cambridge, Massachusetts.

Kiparsky, Paul. 1993. 'An OT Perspective on phonological variation'. Handout from Rutgers Optimality Workshop 1993, also presented at NWAVE 1994, Stanford University. Available at http://www.stanford.edu/~kiparsky/Papers/nwave94.pdf

Kiparsky, Paul. 2003. 'Finnish Noun Inflection', in Satu Manninen and Diane Nelson (eds.), *Generative Approaches to Finnic and Saami Linguistics*, CSLI Publications, Stanford, California. pp. 109-161.

Prince, Alan and Paul Smolensky 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*, Blackwell Publishing, Malden, Massachusetts.

Ringen, Catherine, and Orvokki Heinämäki. 1999. 'Variation in Finnish vowel harmony: An OT account', *Natural Language and Linguistic Theory*, 17, 303-337.

Sadeniemi, Matti. 1949. *Metriikkamme perusteet* [*The Fundamentals of Finnish Metrics*]. Suomalaisen Kirjallisuuden Seura, Helsinki.

Tuomi, Tuomo. 1972. *Suomen kielen käänteissanakirja* [*A Reverse Dictionary of Finnish*], Suomalaisen Kirjallisuuden Seuran Toimituksia 274, Suomalaisen Kirjallisuuden Seura, Helsinki. Available at the University of Helsinki Language Corpus Server http://www.ling.helsinki.fi/uhlcs/.

email: anttila@stanford.edu