

Review of: Arto Anttila (1997): Variation in Finnish phonology and morphology

By Paul Boersma

In this insightful account of Finnish variation data, Arto Anttila convincingly shows that obligatory and variable phonological phenomena can be expressed by a single grammar. Whether such a grammar should have the form that he proposes, namely a PARTIAL ORDER, is a different question. Fortunately, the explicit and detailed presentation allows the reader to replicate Anttila's findings and numbers, which helps us in trying to answer this question.

1. What kind of variation is being modelled?

Anttila uses a written corpus, which shows variation between forms. In general, such variation could be due to variation between lexical forms, to regional, stylistic, or pragmatic factors, to register, to random differences between speakers, and to random variations within speakers. Only in the last case would it be appropriate to regard the corpus variation as generated by a single grammar. The following quote shows why Anttila thinks that variation within the corpus does reflect random variation within speakers:

Native speakers usually report that one variant sounds better than the other while agreeing that both variants are possible. These intuitions are independently confirmed by large corpora where the preferred variant is usually the more frequent one (p. 12).

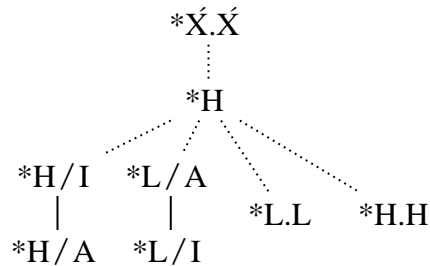
To justify the modelling of corpus frequencies as the result of a single grammar, then, we will have to assume that all speakers share the same grammar, and that the speaker's grammaticality judgments reflect her own production probabilities. With this subject out of the way, I will concentrate on the farther-reaching issues, namely the comparison with other grammar models on points like psychological reality and learnability, which Anttila claims works out to the advantage of his model (pp. 23–29).

2. Anttila's grammar model: partial ordering

Anttila writes his grammars as a kind of strata of subhierarchies, as in (5) in the Summary. However, the class of PARTIALLY ORDERED constraint grammars defined in his book by the properties of "irreflexivity, asymmetry, and transitivity" (p. 5) is larger

than what can be represented by strata of subhierarchies. The actual class is equal to the class of grammars that can be depicted graphically as a dominance hierarchy, so I will graphically represent them as such (the difference will appear crucial below in §4.3). Thus, ranking (5) of the Summary can be depicted as

(1)



In this figure, solid lines represent language-specific rankings, whereas dotted lines represent rankings that Anttila considers universal. We see that $*X.X$ dominates $*H$, and $*H$ dominates $*L.L$, so that by transitivity $*X.X$ dominates $*L.L$. The difference with standard Optimality Theory, however, is that not all rankings have to be specified. Thus, figure (1) does not specify whether $*L.L$ is ranked above $*H/I$, or between $*H/I$ and $*H/A$, or below $*H/A$. This situation may lead to variation in the output for forms in which $*L.L$ conflicts with $*H/I$. For instance, the form *pa.pe.ri.en* violates $*L.L$ (*pe.ri*), and *pa.pe.rei.den* violates $*H/I$ (*rei*), so according to (1) both outcomes are possible.

The **quantitative interpretation** of this variation, according to (7) in the Summary, is as follows. If we consider only the ranking of $*L.L$ with respect to $*H/I$ and $*H/A$, we see that out of the three possible total rankings only one (namely $*L.L \gg *H/I \gg *H/A$) favours the form *pa.pe.rei.den*. We expect, then, the form *pa.pe.rei.den* to occur in one third of the cases, and *pa.pe.ri.en* in two thirds. Adding the influences of $*L/I$ (violated in *ri*) and $*H.H$ (violated in *rei.den*) changes these numbers to 30 and 70 percent, respectively, as shown in (8) in the Summary.

Before touching upon the merits of the general class of partial orderings, I will discuss the simpler subclass of STRATIFIABLE PARTIAL ORDERINGS, which Anttila uses so successfully in chapter 2.

3. Stratifiable partial orderings

3.1. Stratified grammars

The actual grammar that Anttila uses to account for the Finnish *-jen/-iden* choice in genitive plurals, is different from the simplified grammar (1), and this difference will be

crucial in my discussion. The actual grammar can be described as seven STRATA (internally unranked sets) of constraints:

(2)

Stratum 1 (undominated): * $\acute{X}.\acute{X}$ (plus INITIALSTRESS and NOFINALSTRESS)

Stratum 2 (dominated only by the constraints of stratum 1): * \acute{L} , *H

Stratum 3 (dominated by the constraints in strata 1 and 2): *H/I, * \acute{I} , *L.L

Stratum 4: *H/O, * \acute{O} , *L/A, *H.H, * \acute{H} , *X.X

Stratum 5: *H/A, * \acute{A} , *L/O, *A, *L

Stratum 6: *L/I, *O

Stratum 7: *I

In this hierarchy, $\acute{L} \gg * \acute{H}$ militate against stressed light and heavy syllables, *H \gg *L against unstressed heavy and light syllables, * $\acute{I} \gg * \acute{O} \gg * \acute{A}$ against stressed underlyingly high, mid, and low vowels, *A \gg *O \gg *I against unstressed low, mid, and high vowels, and *X.X against adjacent unstressed syllables.

The variable output of grammar (2) can be drawn in a single traditional Optimality-Theoretic tableau, with dotted lines dividing the constraints that are ranked at the same height. Example (3c) in the Summary would become (only three strata are shown):

(3)

paperi-GENPL	2nd str.		3rd stratum			4th stratum					
	* \acute{L}	*H	*H/I	* \acute{I}	*L.L	*H/O	* \acute{O}	*L/A	*H.H	* \acute{H}	*X.X
☞ pá.pe.rèi.den	*	*	*	*		*		*	*	*	
☞ pá.pe.ri.en	*	*			*	*		*			**

There are two winners; which of them wins at evaluation time (i.e. every time a surface form has to be produced), is determined by the coincidental order of the constraints in the third stratum, which will randomly vary between evaluations (none of the constraints in the second stratum has any preference for either form). The form *pá.pe.rèi.den* will win in one third of all cases, namely whenever *L.L happens to be on top of the third stratum, and *pá.pe.ri.en* will win in two thirds, namely when *H/I or * \acute{I} is on top. In a stratified grammar, therefore, determining the production probabilities simply reduces to counting the preferences of the constraints, so that the probabilities will be rational numbers (fractions). Since the decision is made at the level of the third stratum, the cells in the fourth and lower strata can be greyed out, because the violations in these cells can never contribute to determining the winner.

At least, that is the VARIATION INTERPRETATION of tied constraints, which was also defended by Pesetsky (1998: 372): “The output of a set of tied constraints is the union of the outputs of every possible ranking of those constraints”. In the traditional

interpretation of the tie, however (Tesar & Smolensky 1998, p. 241), the violations of all the constraints in a stratum are added up, so *paperien* would be the winner because it has only one violation mark in the third stratum, whereas *papereiden* has two.

3.2. Constraint set

Anttila's large constraint set may arouse the suspicion that several constraints have been included solely for the purpose of probability matching.

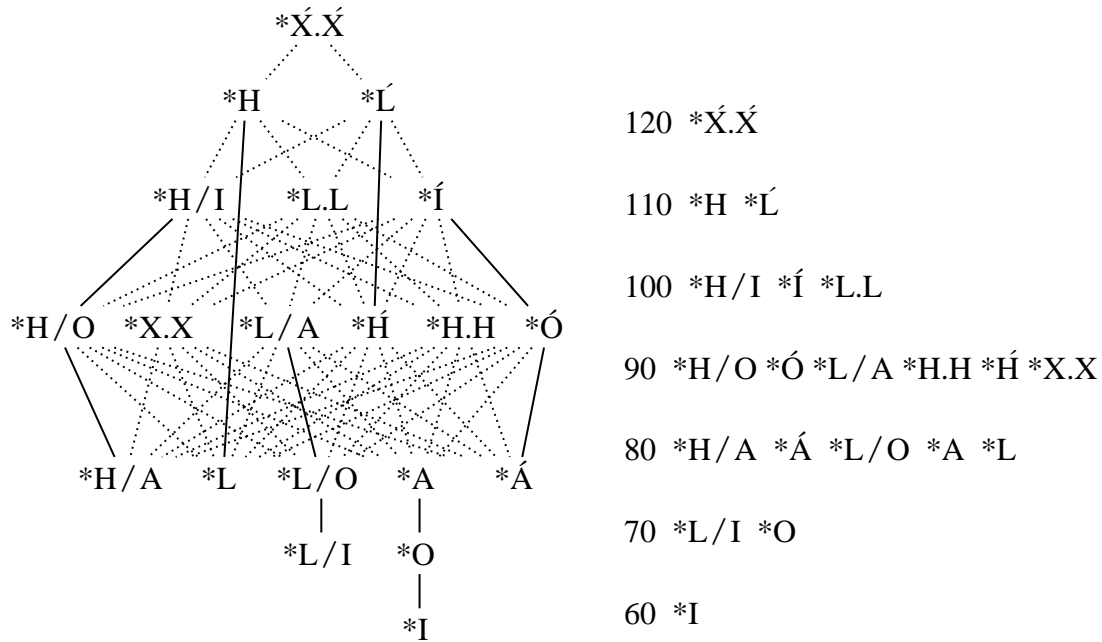
The most controversial seem the constraints that express the anticorrelation between weight and vowel height. If *H/I were left out of the grammar, that would shift *papereiden/paperien* to a 50-50 distribution. Other than for fixing up this ratio in the direction of the attested 2-to-1 distribution, we must believe that Anttila included the weight/height constraints on the basis of observed cross-linguistic tendencies, like the ubiquitous shortening of high vowels. As Anttila convincingly argues in all of his three case studies, the weight/height constraints are at the very basis of many phenomena in Finnish.

Another example is the inclusion of several somewhat perverse sounding constraints like * \acute{H} "no stressed heavy syllables" and * \acute{A} "no stressed low vowels". Including * \acute{H} certainly looks innocent, since it is always dominated by * \acute{L} . However, * \acute{H} crucially comes to support *H.H in making *sáiraalòiden* as bad as *sáiraalojen*, so this perverse constraint does make its influence felt in the output probabilities. The inclusion of perverse constraints is like expressing Prince & Smolensky's (1993) syllable type preferences as the fixed rankings ONSET >> NOONSET and NOCODA >> CODA. This move is certainly a principled and defensible decision: no logically possible constraint is excluded beforehand, and considerations of articulation or perception will predict a cross-linguistically fixed ranking.

3.3. Two more powerful grammar models

Stratified grammars can be described as special cases of partial orders. Grammar (2), for instance, can be represented graphically as in figure (4a) (only the top five strata are shown completely).

(4)



However, stratified grammars can also be described as special cases of continuously ranking grammars with noisy evaluation. In such a grammar (Boersma 1997; 1998, chs. 14–15; Zubritskaya 1997, 143), every constraint has a ranking value along a continuous scale, and at evaluation time a random value (drawn from a Gaussian distribution) is temporarily added to the ranking of each constraint, so that the actual ranking values that are used for determining the winner vary from one output production to the next. Such a grammar can be considered STRATIFIED if all constraint pairs are ranked either at approximately equal height (causing variation with fractional probabilities) or at very different heights (causing categorical, non-variable, behaviour); figure (4b) shows both of these cases.

Both grammar models (partial order and continuous ranking) are more powerful than the simple stratified grammars. There are genuine partial orders, like (1), that cannot be represented as continuous rankings, and there are genuine continuous rankings that cannot be represented as partial orders. It can thus be empirically determined which of the three grammar models is the correct way to describe variation. Since the most restrictive model (stratified grammars) works fine for Anttila’s first example, this model must be our working hypothesis until evidence to the contrary arrives.

3.4. Learnability of stratified grammars

The importance of the stratifiability of a partial constraint ordering lies in its learnability. Whereas no non-exponential learning algorithm has yet been devised for the general problem of partial orders, we are sure that a stratified variation grammar can be learned, because it is a special case of a continuously ranking grammar.

The learning algorithm associated with continuously ranking grammars is quite simple (Boersma 1997; 1998, ch. 15). The learner will repeatedly compare her own output forms with adult forms. If her own form is different from the adult's, she will change her grammar by moving all the constraints that prefer her own form up along the continuous ranking scale (by a small step), and by moving all the constraints that prefer the adult form down along that scale. In this way, the grammar will gradually become more likely to produce adultlike forms.

When applied to the Finnish genitive plurals, the algorithm shifts the 19 constraints, which are initially ranked at the same height, to positions that are quite different from the stratification (2), though *H will still be ranked categorically on top; the output distribution derived from this grammar matches the observed data a little bit better than (2) does, as must be expected on the basis of the added power. If the constraint set is reduced to 13 constraints (by removing the two height/weight hierarchies plus *H and *L), the algorithm still manages to obtain a probability match comparable to the one of (2) (Boersma and Hayes 1999).

3.5. What stratification tells us about the likely powerful grammar model

There is nothing in partial orders that favours stratification. Grammar (4a) does not look like a genuine partial order like (1); rather, it looks like a genuine stratified hierarchy quite artificially forced into the straitjacket of partial ordering. The same, of course, can be said about the artificial continuous ranking in (4b), which looks rather discrete. However, we can show that under certain conditions, the learning algorithm tends to cause constraints to gang up into strata.

If there are lots of evidence for the rankings $A \gg B$ and $A \gg C$, as well as lots of evidence for the rankings $B \gg D$ and $C \gg D$, the algorithm will draw the constraints from one another until A is ranked well above B and C , and D is ranked well below B and C . If the starting point for the learning algorithm is that all the constraints are ranked at the same height, and B and C are never in conflict, it is quite probable that B and C will end up at approximately the same height, so we will probably arrive at the stratified ranking $A \gg \{ B, C \} \gg D$. Now if B and C are conflicting, but only very rarely so, it will take the learner quite a long time to draw apart the rankings of B and C . Thus, B and C will stay in each other's vicinity for quite a long time, and young learners meet many older children whose B and C constraints are not yet categorically ranked in the adult way. Therefore, the average language environment will, in the case of B and C , contain a

lot of variation even if the adult grammar is categorical. This will cause the young learner to acquire the adult ranking of B and C even slower than in the case of a categorical language environment. This again leads to more variation, and the categorical B-C ranking will be lost from the speech community within a few generations. This means that if two constraints are rarely in conflict, languages will often tend to rank them at the same height, so that fractional variation probabilities will arise. In the Finnish case, the relevant constraints conflict only in the case of morphological optionalities, so the condition of relatively rare conflict (as compared with the constraints that determine stress patterns) has been fulfilled.

Within the continuous-ranking model, stratification is expected; within the partial-order model, it is not.

3.6. Partial versus total orders

Anttila maintains that a totally ranked grammar “is the most complex case and presupposes the greatest amount of learning” (p. 29). I want to challenge this, because simplicity depends on your view of the grammar.

In assessing simplicity, Anttila counts the number of ranked pairs. If the grammar is a set of ranked constraint pairs, then partially ordered grammars are simpler than totally ordered grammars. For instance, Anttila’s grammar (4a) needs only 54 ranked pairs in the top five strata, not counting the seven universal rankings. A totally ordered grammar with the same 17 constraints would involve $1/2 \cdot 17 \cdot 16 = 136$ ranked pairs, indeed quite a lot more. An unranked grammar (all constraints in a single stratum), is the simplest grammar, and grammars get more complicated as the number of strata grows.

However, if a grammar is seen not as a list of ranked pairs, but instead as a set of constraints with their properties, a stratified grammar with 17 constraints would need only 17 stratum numbers (one for each constraint), whereas a partially ordered grammar would have to associate a list of dominators with each constraint (e.g., *L.L is ranked above 6 others). Counted in this way, stratified grammars are actually simpler than partially orders.

Moreover, general partial orders need complicated machinery to maintain the transitivity property during learning. Thus, since the hierarchy contains the ranking pairs *L/A >> *L.L and *L.L >> *H, the ranking pair *H >> *L/A should be excluded from consideration, which is not a trivial matter. In a grammar in which each constraint is associated with a stratum number (or with a continuous ranking value, for that matter), such transitivity falls out naturally.

4. Non-stratifiable Anttila grammars

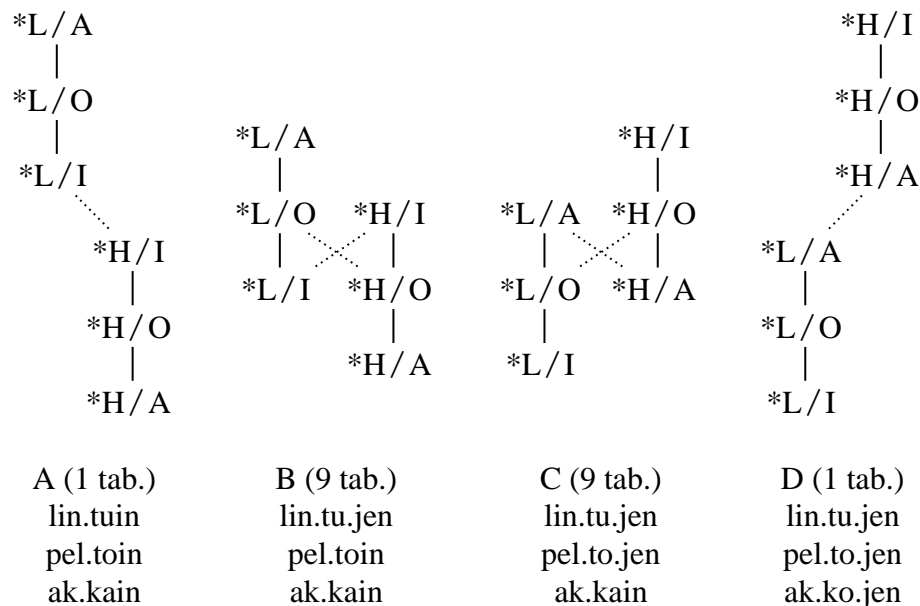
In chapter 3, Anttila gives an account of language change. He adheres to a weak theory of language change, which cannot predict the direction, but according to which the historical

stages only have to be typologically predicted systems, i.e. systems that preserve universal rankings such as $*H/I \gg *H/O \gg *H/A$.

4.1. The change

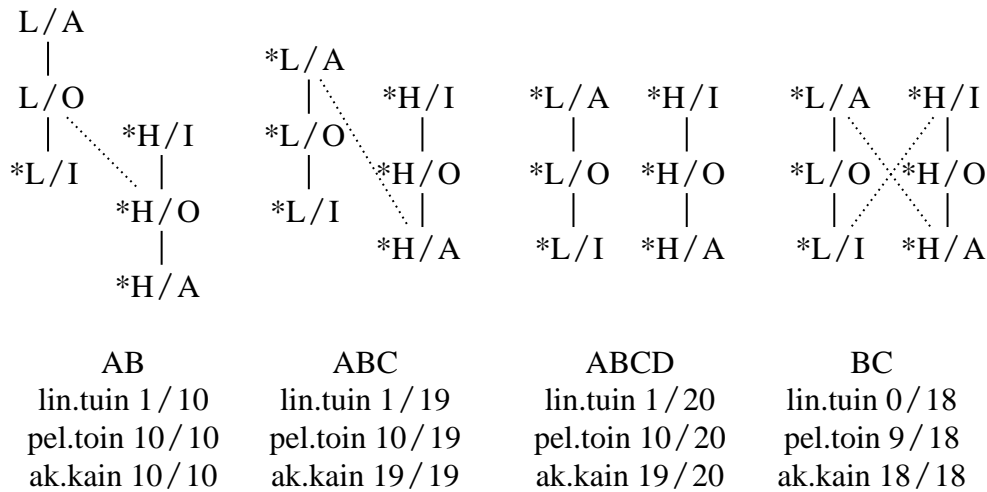
The shift has probably started with an independent sound change, namely the loss of the dental fricative /ð/. Before this change, the genitive plural of /akka/ ‘woman’ was variably /akkaðen/ (short) or /akkoiðen/ (long). After the change, the forms became *ak.kain* (short) and *ak.ko.jen* (long). The first of these has a superheavy final syllable, and it may have been this “problem” that caused a subsequent shift in preference from the short forms to the long forms, starting with the high vowels (*lintu* ‘bird’), and proceeding through the mid vowels (*pelto* ‘field’) to the low vowels (*akka*). Anttila distinguishes the short and long forms on the basis of his familiar height/weight constraints: *ak.kain* violates $*H/A$ (*kain*), whereas *ak.ko.jen* violates $*L/A$ (*ko*). Note that Anttila does not count the $*H/O$ violation in the last syllable of *ak.ko.jen*, which he dispenses with by calling the vowel *e* ‘synchronically epenthetic’ (p. 64, fn. 4). With these constraints, four non-varying grammars are possible:

(5)



Note that including the lowest-ranked of any hierarchy ($*H/A$, $*L/I$) is crucial; otherwise, we would never find *lintuin* or *akkojen*. The typology ranges from all-heavy (A) to all-light (D), with intermediate grammars in which low vowels love heavy syllables and high vowels love light syllables (B and C). Several variation grammars arise from combining grammars adjacent in (5):

(6)



The grammars BCD and CD, not in this figure, are mirror images of ABC and AB.

The well-attested grammars from the 15th century on are more or less in chronological order: AB, ABC ABCD, BCD, CD, D. The intermediate grammars B, C, and BC do not seem to occur; they are exactly the ones with a categorical difference between high and low vowels.

4.2. Why the data point to a single variation grammar

Anttila interprets these results as an argument against a multiple-grammar model. He predicts that grammars like AC and AD, which a multiple-grammar model would allow, are impossible because we cannot regard them as partial orders. To be sure, AD is not attested: no dialect has short and long forms with equal probabilities for high and low vowels; as we see from (6), any possible grammar with variation for all vowel heights must be ABCD, and this grammar will have 5% short forms for high vowels, 50% for mid vowels, and 95% for low vowels.

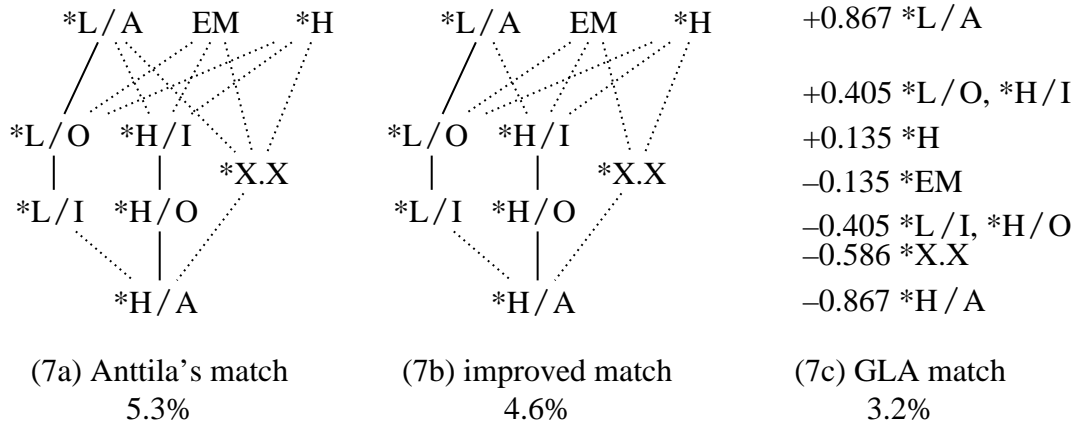
This interpretation is Anttila's central thesis. It still leaves room for each of our three grammar models. However, (6AB) and (6ABC) cannot be represented as stratified grammars, so this leaves partial orders and continuous rankings as the remaining candidates for the modelling of variation in a single grammar.

4.3 Modelling Lönnrot's Finnish

Anttila derives a reasonable probability matching for a 19th-century ABCD-like corpus (the writings of Elias Lönnrot) by adding two old friends to the constraint set: *H (no unstressed heavies) and EM (for ExtraMetricity: no final stresses). Anttila's best match is grammar (7a), a genuine partial order, which is not a completely stratified grammar,

but can still be represented in Anttila’s format, namely as three strata with two subhierarchies in the second stratum.

(7)



The output distribution generated by (7a), which is shown in table (8), matches the observed distribution by a mean absolute error of 5.3 percent; this number is the mean absolute error. In a footnote, Anttila states that “it is possible that an even better one exists”. Indeed, if we sever the dominance of *L/A over *X.X, as in (7b), the predictions improve: the mean absolute error drops to 4.6%. Grammar (7b), though a genuine partial order, is no longer representable as strata with subhierarchies, which may be the cause why Anttila missed it.

We see that small changes to a partial order yield small changes in the expected distribution. This shows that the matches of (7a) and (7b) are hardly evidence for the partial-order hypothesis: it looks as if partial orders sample the distribution space so densely that any distribution can be matched reasonably by a partial order.

Beside partial orders, continuous rankings can match the data as well. Grammar (7c) is a continuous-ranking grammar that, with a noise standard deviation of 1.0, generates a distribution that is only 3.2% removed from the observed distribution. In general, of course, a set of nine continuously rankable constraints (i.e. eight degrees of freedom) is more powerful than a partial order that can generate only 10,080 tableaux, so this slightly better match had to be expected and is no direct evidence for the correctness of either grammar model.

(8) *Matching the probabilities of short forms*

short ~ long	N	observed short	observed percentage	Anttila's match	improved match	GLA match
lin.tuin ~ lin.tu.jen	274	64	23.4%	20.0%	21.9%	22.5%
pel.toin ~ pel.to.jen	145	78	53.8%	45.0%	45.3%	54.6%
ak.kain ~ ak.ko.jen	305	213	69.8%	66.7%	65.6%	70.4%
en.ke.lein ~ en.ke.li.en	58	16	27.6%	25.0%	26.6%	31.1%
va.li.oin ~ va.li.o.jen	15	12	80.0%	75.0%	73.4%	71.1%
a.ka.nain ~ a.ka.no.jen	56	51	91.1%	100.0%	96.9%	87.0%
Mean absolute error				5.3%	4.6%	3.2%
χ^2				∞	13.55	1.95
p ($df = 6$)				0	0.035	0.924

4.4. Evidence for correctness

With a χ^2 test we can compute the probability that a specific proposed grammar model can give rise to a distribution at least as far away from the distribution derived by the proposed model, as the observed distribution is. With six degrees of freedom, as in (8), we expect χ^2 values around 6 if an a priori proposed distribution does underlie the observed data. If the χ^2 value is much smaller, as in the columns “Anttila’s match” and “improved match”, we must reject the hypothesis that the grammar model underlies the data; if the χ^2 value is much greater than 6, as in the column “GLA match”, it becomes likely that the experimenter has matched the model with the data a posteriori.

We note that although Anttila tried a posterior fit of his model to the data, the χ^2 values are low, so that his specific grammar models must be rejected (which does not mean that they cannot be near the truth). Since model (7a) predicts 100% *akanain* forms, the form *akanojen* should not occur. Since this form does occur, the model can never underlie the observed data (χ^2 is infinite, $p = 0$). With the improved partial order (7b), χ^2 drops to a finite, though still high value, giving a probability of 3.5% for this improved model of yielding a distribution as far away from the expected one as we observe. The continuous-ranking grammar (7c) fits the data especially well: χ^2 drops to such a low value, that the data are matched more accurately than they deserve ($p = 92.4\%$), which reveals the effects of doing a posterior fit.



Thus, the observed distribution is a likely outcome for a grammar in the vicinity of (7c), and an unlikely outcome for grammars (7a) and (7b).

5. Errors and other problems


I found few errors of analysis in Anttila's book. On page 73 (and 76), Anttila counts 99 *én.ke.li.en* and 36 *én.ke.li.èn* tableaux; these numbers should be 81 and 54, respectively, but this doesn't influence the result.

In a footnote (p. 18) and an appendix (p. 143), Anttila explains the ungrammaticality of **kla.ri.ne.tei.den* (next to *kla.ri.net.ti.en*) on the basis of the presence of a short geminate *t*, which would make the third syllable heavy. Presumably, Anttila based this on the grammaticality of a prosodically comparable form like *jú.ni.o.rèi.den* (the only H.L.L.H.H form in Anttila's corpus of 7,000 stem types of genitive plurals; occurs 4 times). However, *jú.ni.o.rèi.den* only has to compete with *jú.ni.o.ri.en* (occurs 3 times), which violates a stratum-3 constraint itself, whereas **klá.ri.ne.tèi.den* has to wage an unequal fight against *kla.ri.net.ti.en*, which is phonologically perfect. The forms **kla.ri.ne.ti.en* and **jú.ni.òr.ri.en* are put out of contest by an independent rule of Finnish according to which voiceless plosives, but not sonorants, are geminated between light non-initial syllables; we could tentatively describe this rule by sandwiching the already available structural constraint *L.L between two faithfulness constraints against gemination: *INSERT (length / sonorant) >> *INSERT (length / plosive). The tableaux for the two forms are:

(9)

juniori-GENPL	*INSERT (length / sonorant)	*H	*H/I	*Í	*L.L	*INSERT (length / plosive)
 <i>jú.ni.o.ri.en</i>		*			***	
<i>jú.ni.òr.ri.en</i>	*!	*			*	
 <i>jú.ni.o.rèi.den</i>		*	*	*	**	
<i>jú.ni.òr.rei.den</i>	*!	**	*		*	

(10)

klarineti-GENPL	*INSERT (length / sonorant)	*H	*H/I	*Í	*L.L	*INSERT (length / plosive)
<i>klá.ri.ne.ti.en</i>		*			**!*	
 <i>klá.ri.nèt.ti.en</i>		*			*	*
<i>klá.ri.ne.tèi.den</i>		*	*	*	**!	
<i>klá.ri.nèt.tei.den</i>		**!	*		*	*

Since the correct output forms can be found by evaluating and comparing the surface candidates, the ungrammaticality of **klarineteiden* constitutes no evidence for an underlying geminate *t*. Rather, it provides some evidence for the idea that Anttila's constraint set plays a role not only in the choice between long and short endings, but also in the choice between geminated and ungeminated forms, an idea worth pursuing in the light of the interesting phonology of Finnish gemination.

Another possibly unwanted side effect of the constraints for stress/height anticorrelation is that they predict a different secondary stress placement from the left-to-right rules in the Summary. This occurs in words with heavy third and fourth syllables, if the fourth syllable has a lower vowel than the third. Anttila's corpus contains a couple of these forms, among which *merkityksettömien*, where the constraint set predicts stress on *set* because its vowel is lower than the vowel in *tyk*. Interestingly, the judgments of native Finnish speakers seem to vary on third-and-fourth-heavy words in general, which shows that stress/height anticorrelation may play a role in stress assignment after all.

6. Conclusion

We have seen data that can be represented with stratified grammars (figures 2, 3, 4, 9, 10). For other data, we need a more powerful grammar model, which could be partial ordering (figures 1, 5, 6, 7a, 7b) or continuous ranking (figure 7c). On the basis of what we discussed, we cannot determine which of these two models is correct, though with the present state of formal acquisition models, learnability considerations seem to favour continuous ranking. To decide the issue, many more languages should be investigated, especially with the objective of finding empirical differences between the two models. The virtue of Anttila's exercise, in any case, is that it has showed us the existence and empirical adequacy of a theory that derives variation from a single grammar.

- Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **21**, 43–58.
- Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives [LOT International Series 11]*. Doctoral dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.
- Boersma, P. & B. Hayes (1999). *Empirical tests of the Gradual Constraint-Ranking Learning Algorithm*. Ms. University of Amsterdam and University of California at Los Angeles.
- Pesetsky, D. (1998). Some optimality principles of sentence pronunciation. *Is the best good enough?: optimality and competition in syntax*, edited by P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, & D. Pesetsky, 337–383. Cambridge, Mass.: MIT Press.
- Prince, A. & P. Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers University Center for Cognitive Science Technical Report **2**.
- Tesar, B. & P. Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* **29**, 229–268.
- Zubritskaya, K. (1997). Mechanism of sound change in Optimality Theory. *Language Variation and Change* **9**, 121–148.