

Vowel harmony: learning and generating morphological forms¹

1. Vowel harmony already seems to be a commonplace for a theoretical account. Vowel harmony of Finnish seems to be the most typical example cited in the many general and particular works about this phenomenon. Remarkable that not only single “difficult” cases about Finnish VH are cited from work to work, but also some systematic descriptions of Finnish inflection and word formation in respect of VH are available [Anderson L. 1980]. Also, some psycholinguistic experiments are conducted on how the native speakers tend to choose harmonic suffix variants, when language system provides such an opportunity (disharmonic loanstems) [Levomäki 1972], [Ringen, Heinämäki, 1999].

The system is well-known: Finnish vowels are divided into three classes - ‘back’ (*a, aa, o, oo, u, uu*), ‘front’ (*ä, ää, ö, öö, y, yy*) and ‘neutral’ (*i, ii, e, ee*). We use the terms ‘neutral’, ‘front’ and ‘back’ in inverted commas here and elsewhere to emphasize that they are not phonetic and phonological terms *sensu stricto*, but rather names of functional classes pointing at the phonotactic properties of phonemes. Vowels from the first two sets do not cooccur in native Finnish wordforms, unless the latter are not compounds containing several roots. At the same time, vowels from both sets can cooccur together with ‘neutral’ vowels. All the inflectional and derivational suffixes containing vowels from the first two sets have two allomorphic variants - a ‘front’ and a ‘back’ one. ‘Front’ stems add ‘front’ suffix variants, ‘back’ stems - the ‘back’ ones, respectively. Native stems containing only ‘neutral’ vowels usually (with several exceptions) add front inflectional suffix variants. As for the distribution of derivational suffixes among ‘neutral’ roots, the picture here is more complex (see 4.3.).

But still there’s something to be done about Finnish VH and VH and general. We believe, that some recently elaborated models of language could provide new insights in the theory of vowel harmony and mechanisms of its work in different human language.

2.1. There are two main types of language models. There are *procedural* models, making statements about the *processes* of language use the by native speakers, and *declarative* models, regarding language as an autonomous system and stating the correspondences in it (about these two fundamental types of models see e.g. [Skousen, 2002: 3]).

Among the *procedural* models, the most fruitful we find *usage-based* models, being recently promoted by R.Langacker, M. Barlow, S. Kemmer, K. Johnson, Joan Bybee, R. Skousen etc. (e.g. [Barlow, Kemmer 1999], [Bybee 2001], [Skousen, 2002: 3]). Such theories, to our mind, present psychologically adequate models of language use and change, and also quite realistically explain many human language structural features as resulting from its use by the native speakers.

Among the *declarative* models, we find insightful such model of human language as “Meaning <-> Text” system, elaborated by Russian scholar Igor Mel’chuk. Especially important for us is the morphological module of this system synthesized in his fundamental work “Course of general morphology” [Мельчук, 1997, 1998, 2000, 2001, 2006]. We’ll also cite so-called ‘neostuctural’ model of the language, clearly presented in [Крылов, 2004]. It synthesizes, according to the author’s own words, the achievements of classical structuralism, ‘dynamic’ structuralism (transformational grammar and “Meaning<->Text” system), functionalism, morphological typology, “natural” morphology and phonology, psycholinguistics. Especially important for us in both models is the way of presenting a language as a multi-level system. On each level there are own units, clearly separated from

¹ Supported by Russian Fund of Humanities, grant 06-04-18002e.

the units of other level, and own syntax (the way of combining units within the level). Such a representation can be traced back to the works of post-Bloomfieldians, especially K. Pike with his tagmemics and Ch. Hockett, the school of glossematics created by L. Hjelmslev and H.-J. Uldall (30-40-ies) and later to the stratification theory by S. Lamb and H.A. Gleason and also the theories of S. Bernstein (60-ies). One of the reasons why we prefer Mel'chuk's model is that it is built on the logical definitions of all the terms - this helps to distinguish between the phenomena of different kind and to assemble the phenomena of similar kind in the most explicit way with a minimum of controversy.

Mel'chuk's morphological model in unidimensional and unidirectional (synthesis; meaning -> form), it contains 3 *main levels*: underlying morphological (units are lexemes), surface morphological (morphemes), underlying phonological (phonemes), and 2 *auxiliary levels* between the 2nd and 3rd main levels: level of morphs (morphs) and morphophonological level (phonemes). So this model operates with the main generally known language units, no artificial intermediate units are used. The fewer number of units compared to the number of levels leads Mel'chuk to postulate two types of rules: 1) 'horizontal', acting on the same level (rules of formation=conditions of formal validity=filter rules), that provide well-shaped units for each level; 2) 'vertical', acting between levels (rules of realisation) that put units of *n* level into correspondence with the unit of *n+1* level.

S. Krylov's model is three-dimensional (1) significative, 2) structural, 3) taxonomic relations between units) and bidirectional (meaning->form; form->meaning). So, together there are 6 types of correspondences between units: 1a) expositive (synthetic, transformational, meaning->form); 1b) interpreting (analytical, form->meaning); 2a) divisive (generative, whole->parts); 2b) integrative (selective, parts->whole); 3a) exemplificative (class->subclasses); 3b) classificative (subclasses->class). This model distinguishes more language levels than Mel'chuk's one (though mainly being based on the latter). There are 4 *main levels*: 1) lexogrammatical (=underlying morphological), 2) morphemic (=surface morphological), 3) morphophonemic (=morphological), 4) phonemic (=underlying phonological). The latter is divided into 5 *sublevels* (the levels of abstraction from the concrete phonetic reality), thus presenting very gradual transition from phonemic to phonetic level. Mel'chuk is not dealing with the interaction of phonetic and phonemic levels at all.

The main difference from Mel'chuk's model is that every level and sublevel here has its own units. So on one hand, all the rules act only between levels, they are all 'vertical' - this provides a higher level of symmetry inside the model. On the other hand, such approach equalizes psychologically real main units, like phonemes or morphemes, with auxiliary theoretical constructs like morphophonemes or allophones of several types.

Describing vowel harmony, we'll mainly follow Mel'chuk's model in the sense that we'll use only the main language units - *phonemes* and *morphemes*. We'll also cite allophones - as the main sound types that are distinguished by most of linguists, but do not have distinctive power in the language.

The most important, that both models allow to distinguish quite clearly between different types of information in language systems.

2.2. From the point of view of such multi-level systems, vowel harmony is "a manifold phenomenon, it involves several levels at once, and the question of its substance should be discussed differentially" [Крылов 2004: 201].² Krylov notices further, that vowel harmony implies no alternations on the morphemic and morphophonemic levels. On the level of *morphemes* it's essential only as a boundary signal, distinguishing morpheme combinations inside a wordform from the combinations of wordforms with clitics and other wordforms. On the *morphophonemic* level vowel harmony is a fundamental principle dealing with a

² All citations from works other than written in English are my loose translations.

complementary distribution of harmonic root and affix types. Thus, VH has to do with alternations only on the *phonological* level [ibid.].

For I. Mel'chuk, vowel harmony is essential first of all as a syntactic rule on the level of morphs, a rule of 'agreement' between a dominant morph and recessive morphs [Мельчук 2000: 79]. He also mentions (discussing VH in Hungarian), that distribution of 'front' and 'back' suffix variants is not purely phonological, so the stems should get a respective morphotactic marking [Мельчук 2001: 220].

Many linguists somehow differentiate between the harmony inside the (dominant) morphemes and the harmony across morpheme boundaries (see e.g. [Kiparsky 1973 (1968): 35], [van der Hulst, van de Weijer 1995: 499-500], [Vago 1980: XII]). In multi-level language models these two types of harmony will be presented on different levels.

In the framework of such a multi-level system we could even account for at least *four* types of rules that have to do with vowel harmony:

1) On the levels of *lexemes* and *morphemes* - as a boundary signal (account that can be traced back to N. Trubetzkoy), distinguishing between different types of morphological borders - between lexical morphemes, on one hand, and between grammatical and lexical/other grammatical morphemes.

2) On the level of *morphs* VH is a morphotactic principle in the sense of Mel'chuk. For dominant morphemes, it means their syntactic ability of 'governing' certain allomorphs of recessive morphemes. For recessive morphemes it means the paradigmatic pattern of distribution of their allomorphs between dominant morphemes.

3) On the *morphological* and *phonological* levels, VH means phonotactic patterns of cooccurrence of certain types of vowels inside certain segmental strings, their 'agreement' by any distinctive feature. Sometimes consonants also participate in this 'agreement'. For example, in Votic (one of the Finnic languages) not only two sets of vowels are in complementary distribution in roots and affixes, but also velar *ɫ* and alveolar *l*. Formerly they were allophones, such a situation is preserved in Ingrian, another Finnic language. But after some reduction processes they gained independence, cf. *itm* 'weather' but *silm* 'eye' [Маркыч 2006: 48]. Here it can be seen that the very name 'vowel harmony' is somehow metaphoric.

VH on these two levels is how it has been classically understood. Besides, it's generally referred to as a phonological process. It should be however noticed that we can never formulate the rules of VH in the purely phonological terms, because the borders of VH domain always coincide with any morphological borders. This will be discussed further.

4) On the *phonetic* level VH implies 'agreement' of allophones within a domain limited by certain morphological boundaries. VH is hardly ever understood in such a sense. At the same time, such meaning can be helpful in diachronic perspective, because it's frequently presupposed (for example, for Finno-Ugric languages) that phonological VH has developed from the original allophonic variation inside a wordform (see e.g. [Vihman 1970: 11]). The case of *ɫ/l* variation in Ingrian mentioned above is an example of the phonetic harmony. J. Bybee cites D. Steriade (2000), who "points out that American English flapping of [t] in words derived with *-istic* depends of whether the base word has a flap, even when the phonetic conditions for flapping are met. Thus, *positivistic* (cf. *positive*) does not have a flap, while *fatalistic* (cf. *fatal*) does" [Bybee 2001: 56].

Phonological VH is often referred to as a 'distant progressive assimilation' or as a 'prosody assignment' (in the sense of London prosodic school). Some other features of VH, such as 'phonetic motivatedness', 'bidirectionality', 'unboundness', 'non-optionality' are also cited. Though, as it has been shown e.g. in [Anderson S. 1980], all these criteria are not universal and cause troubles in application to the concrete languages. Lyle Campbell using the example of the word games has shown, that though VH in Finnish doesn't totally undergo any of these phonological criteria, it's still a very powerful and psychologically real phenomenon in

Finnish [Campbell 1980, 1981]. To our mind, usage-based models provide quite sufficient and psychologically real explanation to it.

2.3. Already some of the late generativists in 70-ies, such as P. Linell, J. and M. Ohala, have started to claim that the system of units kept in human brains should not necessarily be maximally economical. Analogy rules are quite productive, so there's a need rather for a vast memory and an ability for a quick search in the lexicon, than for the underlying forms [Fisher-Jørgensen 1975: 291, 295]. These ideas have developed into the usage-based approach that is closely linked also with cognitive linguistics and psycholinguistics. Some summary of the application of usage-based ideas in phonology could be found e.g. in [Bybee 2001]. Following the series of earlier findings in psycholinguistics and citing prototype and exemplar approaches to the language, J. Bybee promotes the idea, that language does not have any specific domain in human brain, but rather uses the same operations that are used elsewhere in mental activity (e.g. abstraction, comparison, analogy transferring, analysis, synthesis etc.). She suggests that language forms are stored in the memory fully specified rather than in their underlying forms. When there's a need for applying any productive rules, for example, for phonological and morphological shaping of newly coming loans, the main mental operation used is analogy, comparison with already present exemplars. "A prototype model of the sort proposed in Rosch (1978) would have the listener store an abstract summary description, or prototype, based on tokens already experienced. Goodness-of-fit ratings would be based on similarity to this prototype. A second possibility is that the listener stores exemplars that are weighed for frequency (Nosofsky 1988). Goodness-of-fit ratings would then be based on similarity to the higher-frequency exemplars" [Bybee 2001: 51]. J. Bybee herself, though, argues more in favour of the exemplar model proposed by K. Johnson. The latter "argues for an exemplar model of speech perception to account for the fact that hearers not only correctly identify words or utterances produced by different speakers, ...but also correctly identify the voice of different speakers they have heard before. In the exemplar model all perceived tokens are categorized and stored, creating categories that directly represent the variation encountered. ...Whatever the limitations on memory, there is certainly enough memory available to allow some version of an exemplar representation; that is, there is no necessity to sort exemplars into prototypes and discard the memory of a particular token. In fact, if tokens of experience were not stored in memory, at least for a while, no prototype could be formed, since categorization depends upon the comparison of multiple individual percepts". [Bybee 2001: 51-52]. So, in such a model (in contrast to the generative one) it's tokens themselves that are 'complicated', but access to them is very simple. According to Bybee (and this idea is not quite new), language system does not change by itself - language change happens only when successive generations of speakers use a language and reanalyse the system. This is why structuralistic competence-based models can't explain adequately the processes of language change - each language structure by itself is so-to-say 'ideal', all the elements are in balanced mutual interrelations, there are no 'better' and 'worse' languages, and each language can express any concept. So, there are in fact no inner structural causes for language change. In spite of what A. Martiné claimed, language systems do not aim anywhere by themselves, do not seek for economy, it's language speakers' brains, that do tend to economize mental efforts.

How then phonetic reanalysis happens? In exemplar framework the answer is quite simple. Phonetic categories are most closely associated with their immediate contexts, particular words, and only via cross-lexical associations - with phonetic categories in other words. "In this model, in which contextually dependent variants constitute prototypes or sets of exemplars themselves, it is easy to see how new phonemes develop: a set of exemplars within a category gradually grows less independent upon context, gradually changes phonetically so as not to be similar to other members of the category, and eventually comes to appear in

contexts where it was not formerly found or to otherwise contrast with sounds with which it was formerly in complementary distribution” [Bybee 2001: 53]. Important point is that “this process of reanalysis can occur very early in the development of a phonological change. ...Restructuring is covert - the speakers change their analysis before the surface forms of the language change (Andersen 1973)” [Bybee 2001: 55]. We also find the similar ideas in the recent works of some Russian scholars. L. Kasatkin, a representative of Moscow phonological school, argues about a “latent” period in the life of a phoneme, stating that “loanwords can manifest regularity that has been covert before. Covert, because there has been no speech evidence for it, though it has already been developed in the language” [Касаткин, 1999: 96]. M. Popov, an adherent of Scherba theories, claims that “a loanword or a neologism (at least, the name of the letter [y]³) is acquired by a native speaker in accordance with his phonological system, so-to-say, it is ‘bolted through a phonological sieve’. ...Occurrence of at least one word with initial [y] in the lexicon already proves the existence of phoneme [y] in the system” [Попов, 2004: 76-77].

Following such an argument, we can finally come to a striking conclusion. If a phoneme, even the most peripheral and distributively limited one, arises as a result of morphologization, first latent, then overt, are there any phonotactic rules, the syntactic rules of phoneme combinations, that would be other than morphophonological ones, i.e. free from any dependence on the morphological boundaries? The answer is presumably no. Any unit, that has already gained a phonemic independence, can *always* potentially occur in an unusual context. From a totally syntagmatic unit, it has become a paradigmatic one. This means, that it can be taken out of its natural segmental string and be used, for example for shaping of new loanwords, or neologisms, or onomatopoeic words, or new composites. It can be even thought in isolation, without any context at all. The idea of alphabet shows this in the most explicit way, presenting graphic phonemic correlates of phonemes as a paradigmatic inventory (though, of course, there’s always a problem of one-to-one letter-phoneme correspondence here).

So, *phonotactic rules* are always potentially violable. Thus, *latently* they are always already *morphophonological rules*. Phonologically, they are more or less stable *tendencies* but no more strict rules. Strict and essentially non-exceptional are only phonetic (allophonic) rules.

Of course, it can be sometimes useful for formal language description to formulate such tendencies in purely phonological terms (although, at least word boundaries are practically always used in such descriptions) - especially, if there are no exceptions to these rules in certain language subsystems (e.g. subsystem of native words, subsystem of disyllabic words, subsystem of nouns etc). But it’s important to understand, that there’s no principal difference between phonological and morphophonological rules.

2.4. Let’s now come back to our original question of vowel harmony. From the previous discussion it can be seen why we treat the phenomenon of vowel harmony as essentially a morphophonological one. Even if we deal with a case, when there are no overt exceptions to VH in the language system (which is very rarely the case, see also [van der Hulst, van de Weijer 1995: 499 etc.]), all the boundaries between harmonizing complexes always coincide with some morphological boundaries (usually between wordforms or lexical morphemes). Non-allophonic VH is a result of morphologization process and usually tends to further morphologization, with more and more exceptions coming into the language system (as it happens e.g. in Finnish and other Finnic languages, such as Votic or Ingrian). That’s why we earlier treated VH on the phonological and morphophonological level together.

At the same time, as we have already said, VH can be psychologically real for the language speakers even on these two levels, not only on the level of morphs. Exemplar-based model

³ in Russian (N.K.)

can also explain this by the analogical processes. When most words are built in a certain way, the tendency of building new words in the same way will be very natural. At the same time, there will always be some percent of vacillation in shaping new words and word forms.

3.1. Finnish word games and shaping of the loanwords in this language show it quite well. L. Campbell points out at some fluctuation in applying VH rules in Finnish word games [Campbell 1981]. Vacillation of similar kind can be observed in the loanwords. For example, there are two normative variants for the word ‘molecular’ - *molekyläärinen* (breaking the harmony principle) and *molekulaarinen* (preserving the harmony) (they are recorded, for example, in [SKP 1990], [Itkonen 2000]). Besides, *molekyläärinen* is also quite common among the speakers, though, at least as far as we know, it’s not recognized by the norm yet. Results of the search for these tokens in Google show that the most frequently used (at least in writing) is *molekulaarinen* (63%), the next one is *molekyläärinen* (25%), and the least frequent is *molekyläärinen* (12%)⁴. Such distribution is quite understandable, if we take two factors in account: normativeness and vowel harmony, with normativeness being stronger than regular system tendency. OT formal account (with a presupposition that constraints are not ranged) could be quite explicit in this case. The first variant is the most frequent, because it’s both normative and harmonic. The second one is normative but totally disharmonic (has a back-front-back vowel pattern). The third one is partly harmonic (right-spreading harmony) but not normative. It seems that many Finnish words containing the element *-ääri-* (of Swedish origin) are likely to demonstrate harmonic vacillation on this particular element. For example, norm allows both *primääri* and *primaari* ‘primary’, *revolutionäärinen* and *revolutionaarinen* ‘revolutionary’, *vulgääri* and *vulgaari* ‘vulgar’. Frequency distribution here in all the cases is close to 50:50. Only for the last word there’s a bit stronger preference of the harmonic ‘back’ variant, about 68%. It’s probably connected with a primarily stressed back vowel in this word, what will be discussed further.

Vacillation in language use not yet recognized by the norm can be seen, for example, in the stem, originating from a French surname ‘Pasteur’. The only normative are the forms *pastöroida* ‘pasteurize’, *pastörinti* ‘pasteurization’. They both are totally disharmonic, so we found in Google about 10-12% of forms *pastoroida* and 5-10% of *pastorointi*, which are non-normative, but harmonic. Probably, this percentage will be even higher in speech, because such sequences of vowels are really very untypical for Finnish and thus difficult for the native speakers to pronounce. The same situation is observed with the famous example *olympialaiset* that can be sometimes pronounced as *olumpialaiset*, that was used by R. Skousen to prove the productiveness of VH within the morphemes [Campbell 1981: 157], see also [Skousen 1973]). In Google we find up to 10% of *olumpialaiset*.

At the same time, it should be stated that such fluctuation showing activeness of VH tendency on the phonological and morphonological levels is not so widespread. The majority of loan stems do not follow VH, if only by chance. A tendency to adjust loanwords according to the principles of VH reveals itself only in the utmost cases, when such words produce real pronunciation difficulties for the native speakers (like vowel sequence *a-ö-o* in *pastöroida*). Otherwise VH is not followed.

3.2. Much more interesting case is how VH works on the level of morphs, i.e. which are the principles of adding either ‘front’ or ‘back’ suffix variants (or both) to disharmonic loan stems? I.e. how the structure of stems affects their syntactic (morphotactic) behaviour? This question, as we have already mentioned in the beginning, from the psycholinguistic point of

⁴ These are results for the joint search for word forms in Nominative and Genitive. We searched for Nominatives only for the words showing vacillation in VH on the phonemic level, for Partitives only for the words, showing VH vacillation on the morphophonemic level and for both forms when a word showed VH vacillation on both levels. Our percentage presented in this paper is an arithmetical mean between ‘the most relevant results’ and ‘all the results’ shown by Google.

view has been more or less systematically investigated in [Levomäki 1972; Ringen, Heinämäki, 1999]. We have also conducted our own small research on the words mentioned in these two articles via Google (searching for the wordforms in Partitive Sg, that has *-a/-ä-* suffix added to the stem). Various strategies are found here: from *syntaksi* ‘syntax’ always taking back suffix variants (e.g. in Partitive: *syntaksi-a*), via *analyysi* ‘analysis’, where the statistic distribution of speakers’ preferences is about 50:50 (i.e. both *analyysi-a* and *analyysi-ä* occur with same frequency), towards *miljonääri* ‘millionaire’ always adding front suffix allomorphes (e.g. *miljonääri-ä*). But these are only the main points on the scale - we find a very wide spectrum of statistic distributions of speakers’ preferences for such words.

Some principles leading the behaviour of native speakers has already been proposed by these are other researchers. The most obvious are the two of them:

1) if a disharmonic stem ends in a ‘back’ vowel or a diphthong including ‘back’ vowel, then it adds only ‘back’ suffix variants (e.g. stems *pastöroi-* and *lyyra-* ‘lyre’);

2) behaviour of the vowels *y* and *yy* is different from the behaviour of other front ‘non-neutral’ vowels *ä, ää, ö, öö* in the sense that the former ones are ‘weaker’ in their ‘front harmony power’. There has been such an opinion among linguists, that *y, yy* behave like ‘neutral’ vowels in the loanwords. In [Levomäki 1972] it has been proved, that it’s not quite the case. When there’s a loan stem, that ends on one or two ‘neutral’ vowels with preceding *ä, ää, ö, öö* (like *miljonääri-* or *interiööri-* ‘interior’), native speakers strongly prefer ‘front’ inflectional suffixes over ‘back’ ones. When a vowel preceding final ‘neutral’ stem vowel(s) is ‘back’ (like in *syntaksi, turisti* ‘tourist’, *artikkeli* ‘article’), ‘back’ suffix variants are strongly preferred. But the peculiarity of vowels *y, yy* manifests itself in that loan stems, having them before final neutral vowels (like *analyysi, hieroglyfi* ‘hieroglyph’), usually vacillate between ‘front’ and ‘back’ suffix allomorphs without strong preference of either. Respectively, norm allows both types of inflection for such stems, but only one type for the both previous cases (‘front’ in the former case and ‘back’ in the latter).

At the same time, researchers somehow tend to go further, than just stating this obvious peculiarity of *y, yy*, and try and give a very detailed classification of vowels in terms of their ‘harmonic power’, the ability to influence on choosing either front or back suffix allomorphs. For example, L. Anderson presents a typological scale, based on his investigation of Finnish, Hungarian, Turkish and Mongolian: from the ‘most ‘neutral’ towards ‘harmonically stongest’: *i > e > u > a > o* [Anderson L. 1980: 274]. In [Ringen, Heinämäki, 1999] the authors claim such a ‘sonority scale’ for Finnish (from the most to the least sonorous vowel): *ää, aa > öö, oo, ee > uu > ä, a > ö, o, e > ii, yy > i, y, u*.

3.3. Let’s now look at the grounds of these very detailed classifications. As for Anderson’s proposal, his classification is based on his research on the whole derivational and inflectional systems of both native and loanwords in these languages. Moreover, the investigation of the *derivational* system of *native* words is the main goal of his research. We’ll discuss the significance of this fact further.

As for C. Ringen and O. Heinämäki’s account, it focuses on the contrary on the *inflectional* system of *loanwords* in Finnish. Remarkable is that they don’t really give any solid explicit grounds for such a scale, they just ‘assume’ and ‘suggest’ it. In fact, from the OT perspective it’s quite understandable - you have a body of empiric data (in this case, statistic distributions of speaker’s preferences in choosing either ‘front’ or ‘back’ suffixes for certain loanwords of different structure), and your goal is to find such an arrangement of different constraints, which would fit this data body best. If the joint work of your constraints gives you back your empiric data, then it’s considered to be effective and doesn’t require any further justification. Efficiency is the main criterion. The problem here is that if you don’t have any fundamental grounds for your classification, and especially if you classification is a very detailed one and the data are statistic, the risk is high, that if you add any new data to your body, they will no

more fit the constraints combination. When we tried via Google all the words from Ringen&Heinämäki's article, we got generally the same results as them. But when we further tried other words from Levomäki's list and also added some new words, then we got of course some controversy with their OT account. For example, a word *molekyyli* 'molecule' for some reason has a strong preference of 'front' suffixes over the 'back' ones - about 90%. Levomäki also gives 89,5% of 'front' variant preference for this word. This word is of *analyysi* type, it can hardly be analysed as a composite word (unlike e.g. *stereotyyppi* 'stereotype', that gets about 98% of 'front' suffix preference). Norm allows both suffix variants here (while only 'front' variant for *stereotyyppi*). Maybe some other factors are also in play here. For example, maybe adjective variants derived from this word that has been considered above somehow influence the base word. Also a word *trotyyli* 'trotyl' (of the same *analyysi* type) seems deviant, getting as much as 70% of 'front' harmony. A word of similar structure *marttyyri* 'martyr' having 'more sonorous' (in terms of the article authors) primarily stressed vowel than *trotyyli* gets a regular distribution close to 50:50. *Trotyyli* with its 3 syllables does not even get a secondary stress and can be even less than *molekyyli* analysed as a compound.

There are altogether four constraints regulating the preference of suffix variants for the loanwords, proposed in [Ringen, Heinämäki, 1999]:

1) NO-INTERVENING [-BACK]

No vowel intervenes between a [-back] feature and the right edge of the word;

2) PRIMARY STRESS

All vowels have the same specification for [±back] as does the vowel with primary stress;

3) SECONDARY STRESS

All vowels have the same specification for [±back] as does the vowel with secondary stress;

4) SONORITY

All vowels have the same specification for [±back] as does the most sonorous vowel of the root

[Ringen, Heinämäki, 1999: 320].

The authors suggest that Primary Stress constraint is ranked first, but all the other constraints are unranked, and being combined in different ways by native speakers, create the observed statistic usage variation. Thus, the authors support the proposal of P. Kiparsky (1993) and A. Anttila (1997), that some constraints are not ranked in the grammar [Ringen, Heinämäki, 1999: 331].

3.4. As for the sonority scale, our observations showed that there's significant statistic difference only between three main groups of (phonetically) front vowels already mentioned before.

Practically all the words, having *ö, öö, ä, ää* before final stem 'neutral' vowels, show strong preference of 'front' suffixes, about 98-100%. These are such words, as *miljonääri, primääri, afääri* 'affair', *revolutionäärinen, hydrosfääri* 'hydrosphere', *interiööri, kasööri* 'cashier', *amatööri* 'amateur', *sutenööri* 'pimp', *jonglööri* 'juggler', *konduktööri* 'conductor'. The only word that showed just 90% 'front' suffix preference was *vulgääri*. It somehow correlates with the fact that *vulgääri* wordforms, both with 'front' and 'back' suffixes, are still less frequent than *vulgaari* wordforms. It could be probably influenced by primarily stressed back vowel in *vulgaari*. To compare, we can say that *revolutionäärinen* and *primääri* are on the contrary more frequent than their harmonic back stem counterparts (see 2.5.). It should be also mentioned, that all the words from the above listed series, that have (phonetically) front primarily stressed vowel, showed 100% of 'front' suffixes, so no vacillation at all. Among other words, 3-syllable *afääri* and *kasööri* get 100% of 'front' suffixes; others show some 1-3% vacillation. Though all these difference are not so significant to prove reliably any hypothesis.

As for the group of words having *y, yy* before the final stem ‘neutral’ vowels, it somehow splits into two subgroups, though a border between them is not so clear. First one consists of words having a distribution of ‘front’ and ‘back’ suffixes close to 50:50 (40-60% of ‘front’ suffixes). Second one - of words having more than 75% of ‘front’ suffixes up to 100% (but the majority of words lying within 80-86%). First group includes words *synonyymi* ‘synonyme’, *analyysi*, *apokryfi* ‘apocryphal story’, *marttyyri*, *psykoanalyysi* ‘psychoanalysis’. Second group includes words *polymeeri* ‘polymers’, *anhydridi* ‘anhydride’, *asymmetrinen* ‘asymmetrical’, *hieroglyfi*, *manikyyri* ‘manicure’, *manikyristi* ‘manicurist’, *elektrolyysi* ‘electrolysis’, *metafyysinen* ‘metaphysical’. On the bottom of this group is *trotyyli*, on the top (and in fact statistically already belonging to the *ää, ä, öö, ö* group) - *molekyyli*, *aldehydi* ‘aldehyde’, *stereotyyppi* and *antipyriini* ‘antipyrine’.

The probable tendency dividing these groups has morphological roots. It has already been supposed in several works, starting from M. Sadeniemi (1949) that some loanwords can be analysed by native speakers as composites [Ringen, Heinämäki, 1999: 310-311]. It’s obviously the case, when both parts of a loanword can exist separately. For example, together with *stereotyyppi* Finnish language has borrowed *stereo-* and *tyyppi*. So, this word is a certain compound, no wonder why it gets 98,5% of front suffixes - only the part *tyyppi* is decisive here in the choose of harmony. The same can be said about *antipyriini* - a word *pyriini* also exists, and loan affix *anti-* can be found in other loanwords, for example, *antimilitarismi* ‘antimilitarism’, *antiteesi* ‘antithesis’.

In fact, the similar observation can be made for most words from the 2nd *y, yy*-subgroup. For many of them, at least one part is fully meaningful: *elektro-*, *hydridi*, *symmetrinen*, *fyysinen* are separate roots, *-lyysi-*, *-nyymi-*, *-glyfi-*, *poly-*, *anti-*, *-hydi-*, *-leeni-* are derivational loan affixes (or affixoids) that occur in more than one loanword. So, *elektrolyysi* apparently gets more ‘front’ suffix variants than *analyysi*, because in the former both parts possess some kind of lexical independence. The former word having the structure ‘root + derivational suffix’ is much more morphologically *analogical* to other Finnish words than the latter. *Psykoanalyysi* is apparently associated with *analyysi* and the first morphological part of the word is also disharmonic, that’s why this word has a distribution of 40% ‘front’ suffixes - close to *analyysi* but lower.

But of course no strict borders exist here, as elsewhere in the language. For example, as we already said, it’s difficult to explain why *molekyyli* gets so high percentage of ‘front’ suffixes. *Manikyyri* can also be hardly divided into morphological components, though it gets 80% of ‘front’ suffixes. Important is to find fundamental language tendencies underlying the most significant statistic differences.

4.1. On the whole, the variety of factors can influence the preferences of the speakers, among them:

1) Above mentioned morphological structure of the word, morphological independence of its parts from each other;

2) Foot structure of the words, defining stress pattern (the occurrence of secondary stress in the words having more than 3 syllables) - phonological independence of word parts from each other.

These two factors seem to be the most important - they have something to do with *separating* of the word, increasing the independence of its parts and making inner borders in the wordform stronger. As we can see from the distribution and judge from the general linguistic principles and principles of VH ‘spreading’, in particular, the first, *morphological*, constraint is more important than phonological one.

The importance of morphological factor can be also observed within the corpus of disharmonic words containing only ‘neutral’ and ‘back’ vowels. The words, the second parts of which is apparently associated with words, containing only neutral vowels, could get up to

100% of 'front' suffix choice: *partenogeneettinen* 'parthenogenetic' inflects like *geneettinen* 'genetic', *heterogeeninen* 'heterogeneous' as *geeninen* 'genic', *kilometri* 'kilometre' - as *metri* 'metre', as well as *heksametri* 'hexameter'. In *barometri* 'barometer' the element *-metri* is also separable as a suffixoid, by association with other words ending in *-metri*, though the part *baro-* can't be used independently. So the norm would prescribe not to try to separate such a word (see a note about *-geeni* and *-metri* words on the Language Committee homepage http://www.kotus.fi/kielitoimisto/kielijutut/2005_2_omaajavierasta.shtml). Obviously, *-tehti* and *-melli* in *arkkitehti* and *karamelli* are also analysed as suffixoids by native speakers, so these words get quite a high percent of 'front' suffix variants. It is specially discussed on the same homepage, that these words are not compounds.

Probably, also the following factors, mentioned in [Ringen, Heinämäki, 1999] can somehow participate in the choice of native speakers:

- 2) Is primarily stressed first vowel (phonetically) back of front one;
- 3) Occurrence of long (phonetically) front vowels in the word, especially in non-first syllables.

But again, no obvious regularities could be observed here.

4.2. Some authors (P. Kiparsky (1981), R. Välimaa-Blum (1987), D. Steriade (1987)) have suggested that there's a stylistic difference in using 'front' and 'back' suffix variants. In [Ringen, Heinämäki, 1999] this thesis is disputed. But we have also got evidence from the native speakers, that there could be some stylistic difference. One speaker who has to do with language correction and normalization told us, that in case of *yy*, *y* words, where both suffix variants are normatively allowed, he would probably pronounce 'front' variant, because it's 'easier and more natural', but he would always write 'back' variant, because it somehow 'more learned'. He has also noticed that the peculiarity of *y*, *yy* can possibly arise from the orthography, the way how they look like in writing. The letter *y* is not specially marked by diacritics, as *ö* and *ä* are. So e.g. *analyysi* looks more like the original *analysis*, not so "Finnish-like" than e.g. *interiööri* (from to the original French *interieur*). So to say, if a letter doesn't have graphic marking of the 'front' class, it also has less 'power' in causing 'front' suffix variants.

4.3. Let's now come back to the question we have for some time abandoned - the significance of the fact, that L. Anderson builds his detailed vowel typology on the analysis of the *derivational* system of *native* words. And also - what is the principal difference of his research from the psycholinguistic one, conducted by M. Levomäki and C. Ringen and O. Heinämäki? This question leads us back to the exemplar-based approach. The derivational system of native words as well as their inflection from the exemplar perspective is rather learned than generated by the rules (at least in Finnish). That's why it's quite fruitless to speak about generative rules operating in such a system on the *synchronic* level. It's usually quite pointless to say that such and such combination of vowels and consonants in the root triggers such and such harmony in the stem vowel or in derivative suffixes or that such and such vowels 'block' the harmony, because these tendencies are not living in the language. As we have said in the beginning, VH is already a morphophonological phenomenon, at least in Finnish, so these purely phonological terms seem a bit inadequate to its description. We can analyse Finnish VH in such terms only *diachronically*.

But the situation observed in the formation of loanwords (both their phonological structure and morphotactic properties of the stems) is of different sort. Here we can really see living generative powers actively participating in the speech and writing behaviour of the native speakers. In the native words, that are being learned by the speakers, the variation observed is a distribution of either front or back derivational suffixes among *different* words. So, as it has already been mentioned, there are no VH *alternations* on the morphophonemic level here. The complexity of this distributional pattern is a result of different phonological tendencies

having worked in Finnish in different historical periods of time. It is very well summarized in two tables and two appendixes in [Anderson 1980: 276-277, 332-340]. The basic tendencies that can be seen from these tables are the following:

1) Among phonetically front vowels, only *e* and *i* can occur in native roots together with phonetically back vowels. This is a tendency defining the phonological structure of the roots. As for their morphotactic properties, we learn that among roots containing only phonetically front vowels, roots containing only *e* and *i* add in some cases suffixes with back vowels. There are not so many exceptions for inflectional suffixes here, but the pattern of derivational suffix distribution is quite elaborated. These properties of *i* and *e* made linguists to put them into the functional class of ‘neutral’ vowels.

In general, one can see from Anderson’s tables, that stems/roots where *e* is ‘decisive’ for harmony have some advantage over stems/roots with *i*. It is valid both for phonological structure of the roots and their morphotactic properties. This naturally makes Anderson to place *i* lower than *e* on the scale of ‘harmonic power’.

As for *y*, it doesn’t differ very much from the other ‘front’ class vowels. It behaves the same way in the inflection and word formation of the native stems. The only difference that Anderson mentions (after R. Austerlitz) is that in loan stems with *y*, *yy* containing derivational suffixes *-iikka-* and *-ikko-* got back vowel *-a-* in these suffixes: *lyriikka* ‘lyrical poetry’, *lyyrikko* ‘lyricist’ [Anderson L. 1980: 295]. The problem with morphological analysis of these words is to what extent these segments can be separated from the root and called ‘suffixes’? For sure, they are suffixoids, because there’s a whole class of categorically similar words containing this element, but it’s hard to say about the degree of their morphological independence. Hard to say, to what extent *lyriikka* and *lyyrikko* are associated with *lyyra* ‘lyre’, can they be considered derivatives from the latter or not. Besides, the element *-iikka-* does not occur in the words where ‘front’ harmony would be required (in stems containing *ö*, *öö*, *ä*, *ää*). It’s also not clear to what extent we could associate this *-ikko-* with a suffix *-ikko-*/*-ikkö-* occurring in the native stems (e.g. *säle* ‘plank’ -> *säleikkö* ‘Venetian blind’). Anyway, whatever analysis is taken for granted, such stems with *y*, *yy* would differ from stems with other ‘front’ vowels. Both such a class is relatively small and consists only of recent loans. It means, that such a peculiarity of *y*, *yy* have manifested itself quite recently and was not actually caused by the interrelations already existing in the system. It could rather be explained from the ongoing process of VH morphologization that leads to the popping up of more and more harmony ‘exceptions’.

2) Native vocabulary of Finnish demonstrates us also that VH is directional in this language. It’s difficult to say exactly how it all worked in Proto-Finnic, when it was probably still allophonic tendency, but at least it morphologized as a root-controlled system. In such a system there are dominant morphemes that direct the choice of either ‘front’ or ‘back’ suffix variant.

3) VH is sensitive to the type of morphological and morphophonological borders inside and between the wordforms. Anderson gives a very detailed scale of morpheme ‘strengths’ in their ‘resistance’ to ‘front’ harmony:

second parts of compounds > derivational suffixes with [grave] consonants + /o/ (this group is further divided into 5 subgroups of particular suffixes > vowel-initial suffixes > meaning-“transparent” derivatives > true inflections.

It can be seen, that the more ‘lexical’ the border is, the more ‘reluctantly’ VH spreads across it. And on the contrary, the more ‘grammatical’ the border is, the more ‘transparent’ it is to VH. Lexical borders are harmonically ‘strong’, grammatical borders are harmonically ‘weak’. Besides, a phonological structure of the morphemes has some impact on the strength of the border.

4.4. Now, this is the system that native speakers learn. The distribution of suffix variants here is static, practically nothing is actually generated here. A very small dynamic fluctuation inside the norm is observed only in several comparatives, like *ille-mma-lla* / *ille-mmä-llä* 'later on in the evening' [Anderson L. 1980: 281]. On the contrary, the inflection and sometimes the phonological structure of recent loans in Finnish (as in many other languages) is a subject to active language processes, to dynamic generation. Norm is being corrected here all the time, and these changes mainly follow the creative processes in spontaneous language use. The same can be said about language games, described by L. Campbell. From our point of view, these two domains of the language, static one and dynamic one, are principally different and shouldn't be mixed. If one wants to understand how and to what extent VH works in Finnish nowadays as a productive morphophonological tendency, he should first of all pay attention to the two latter subsystems of language.

It's very interesting to compare VH effects in the already given system of native words and in the system of loanwords and language games. Being of different nature, they are, of course, interrelated, the former being the basis for the latter. It can also be well explained within the exemplar and prototype framework. Learned exemplars create prototypes that are later used in generative language processes. It's curious to see how the static rules of original system transform into active dynamic rules, which properties of the former are saved and which ones are lost.

In general, we see, that in dynamic perspective VH on the levels of morphs and morphemes is more important than on the phonological and morphophonological ones.

On the two latter levels we can observe the adjustment of the most structurally untypical loanwords (with a vowel sequence 'back'-'front'-'back' in the stem) to the more habitual patterns (see 3.1.). Let's notice that the principle of VH directionality, depicted in the static form in the system of natural words is actually active in Finnish. There are two ways of adjusting such words as e.g. *molekyläärinen*: 1) either a disharmonic front vowel in the middle is changed into the back one (*molekulaarinen*); or 2) front harmony 'spreads' to the right end of the stem (*molekyläärinen*). So, here we can observe some isomorphism of static and dynamic tendencies.

On the level of morphemes (and lexemes) VH indeed serves as a boundary signal helping native speakers to separate wordforms in speech, as it has been shown in [Suomi, McQueen, Cutler 1997]. By the way, here it's interesting to notice also some kind of 'directionality'. VH, at least in Finnic languages, getting further morphologized, starts to 'degenerate' from the end of the word, from the 'recessive' morphemes. More and more suffixes start to change their morphotactic properties getting 'back' harmony instead of the 'front' one, first derivational (more 'lexical'), then inflectional (for example, Commitative case suffix *-ka-* in Votic). The process of harmony break-up seems to be closely connected with a phonetic quantitative reduction further turning into a qualitative one. For example, in Ingrian (that has Finnish type VH) suffixes from certain non-first syllables often get merged in pronunciation because of a strong reduction in these syllables. The utmost variant of harmony loss we can see, for example, in Estonian, where in native stems 'front' vowels occur only in the first syllable of the root.

So, due to such a tendency, the vowel sequence 'front'-'back' in a Finnish wordform is much more common than a sequence 'back'-'front'. This can well explain the fact observed in the above-mentioned article, that the latter sequence serves as a more clear signal of a word boundary (or a boundary between roots) in speech perception.

On the level of morphs VH seems to be the most powerful - it's a dynamic mechanism, regulating the choice of 'back' and 'front' suffix variants for the disharmonic loan stems. As it has been said, two main tendencies are observed here:

1) Three main groups of vowels are clearly distinguished by their 'harmonic power' (from the 'strongest' to the 'weakest' ones): *ää, ä, öö, ö* > *yy, y* > *ee, e, ii, i*. It's remarkable that this tendency is on one hand isomorphic to the static one observed in the native stems, but on the other hand not totally identical to latter. While on the basis of the native vocabulary it's possible to built a very detailed hierarchy of vowels (exactly what L. Anderson does), in the dynamic perspective differences between *ä* and *ö* and *e* and *i*, as well as quantity differences between the vowels of the same quality, turn out to be statistically insignificant. Only the main proportions and contrasts become productive. At the same time, the difference of *y, yy* from both other groups of (phonetically) front vowels somehow seems to be a recent innovation, not quite supported by the native vocabulary. It's an active power on the synchronic level with a short past. Maybe, this peculiarity has appeared partly due to the character of the letter itself, as we have discussed above. All these loanwords belong to the learned vocabulary, so their primary state is rather to be written and read, than pronounced and heard.

2) The second tendency is the role of morphological boundaries of different type for the VH. Again, only the main principle is preserved as an active tendency: borders between wordforms and roots restrict the domains of VH. The clearer and the more 'lexical' is the border inside a word, the higher are the chances that only the last part of the word will be taken into account in the choice of suffix variant. There's no such a detailed hierarchy of suffixes and parts of the compounds, as L. Anderson describes it for the native stems, but again the isomorphism of static and dynamic tendencies is preserved.

So, we tried to show that there are quite solid grounds to distinguish between the static distribution of certain forms in the part of the language system that is learned by the native speakers and the dynamic generative powers involved in the active creative processes on the synchronic level. In case of Finnish VH we see, that distribution of 'front' and 'back' suffixes among different lexemes, depicting the traces of previous phonological processes, forms the basis for the generative strategies on synchronic level. The latter ones are isomorphic to the former ones, but not so elaborated - human brains are likely to activate only the major system contrasts in their generative analogical work. The interesting difference between these two subsystems is that static vacillation on the level of morphemes (structure of the morpheme determines the only suffix variant for each word, though for the words of the similar structure the choice can be different) transforms into the dynamic vacillation on the level of morphs (both suffix variants are possible for the same word, and here the structure of the word determines the statistic percentage of either choice).

Bibliography

Anderson L. 1980 - Anderson L. B. Using asymmetrical and gradient data in the study of vowel harmony / Vago 1980, pp. 271-340.

Anderson S. 1980 - Anderson S. R. Problems and perspectives in the description of vowel harmony // Vago 1980, pp. 1-48.

Barlow, Kemmer 1999 - Usage-based models of language / ed. by M. Barlow, S. Kemmer. Standford, Cal.: CSLI Publications.

Bybee 2001 - Phonology and Language Use. Cambridge University Press.

Campbell 1980 - Campbell L. The psychological and sociological reality of Finnish vowel harmony // Vago 1980, pp. 245-270.

Campbell 1981 - L. Campbell. Generative phonology vs. Finnish phonology: retrospect and prospect // Phonology in the 1980's / ed. by D.L. Goyvaerts. Ghent/Belgium.

Fisher-Jørgensen 1975 - Fisher-Jørgensen E. Trends in Phonological Theory. A Historical Introduction. Copenhagen: Akademisk Forlag.

Itkonen 2000 - T. Itkonen. Uusi kieliopas. Helsinki: Kustannusosakeyhtiö Tammi.

Kiparsky 1973 (1968) - P. Kiparsky. How abstract is phonology? // Three dimensions of linguistic theory / ed. by Osamu Fujimura. Tokyo Institute for Advanced Studies of Language, pp. 5-56.

Levomäki 1972 - M. Levomäki. Vierasperäisten sanojen suffiksaali vokaalisointu // Virittäjä 76, s.254-261.

Ringen, Heinämäki, 1999 - C. O. Ringen, O. Heinämäki. Variation in Finnish vowel harmony: an OT account // Natural Language and Linguistic Theory 17, pp. 303-337.

Skousen 1973 - R. Skousen. Finnish vowel harmony: rules and conditions // Issues in Phonological Theory. Proceedings of the Urbana Conference on Phonology / ed. by M. Kenstowicz, Ch. W. Kisseberth. The Hague-Paris: Mouton, pp. 118-129.

Skousen 2002 - Analogical modelling. An exemplar-based approach to language / ed. by R. Skousen, D. Lonsdale, D. B. Parkinson. Amsterdam/Philadelphia: John Benjamins Publishing Company.

SKP 1990 - Suomen kielen perussanakirja. Helsinki: Kotimaisten kielten tutkimuskeskus.

Suomi, McQueen, Cutler 1997 - K. Suomi, J. M. McQueen, A. Cutler. Vowel harmony and speech segmentation in Finnish // Journal of Memory and Language, 36, pp. 422-444.

Vago 1980 - Issues in Vowel Harmony. Proceedings of the CUNY Linguistic Conference on Vowel Harmony, 14th May 1977 / ed. by R.M. Vago. Amsterdam / John Benjamins B.V.

van der Hulst, van de Weijer 1995 - H. van der Hulst, J. van de Weijer. Vowel Harmony // The Handbook of Phonological Theory / ed. by J. A. Goldsmith. Oxford, pp. 495-534.

Vihman 1970 - M. Vihman. Vowel harmony in Proto-Finnic: The Livonian Evidence. Baltic Conference. MS.

Касаткин, 1999 - Касаткин Л.Л. Современная русская диалектная и литературная фонетика как источник для истории русского языка. М.: Наука-Школа «ЯРК», 1999.

Крылов, 2004 - Крылов С.А. Теоретическая грамматика современного монгольского языка и смежные проблемы общей лингвистики. Ч.1. Морфемика, морфонология, элементы фонологической трансформаторики (в аспекте общей теории морфологических и морфонологических моделей). М.: Вост. лит.

Маркус 2006 - Маркус Е.Б. Типология морфемного варьирования (на материале морфонологических систем говоров водского языка). Дисс. на соиск. уч. степ. канд. фил. наук. Институт языкознания РАН. М. 2006. Рукопись.

Мельчук 1997, 1998, 2000, 2001, 2006 - Мельчук И.А. Курс общей морфологии. Т.1, Т.2, Т.3, Т.4. Москва-Вена.

Попов 2004 - Попов М.Б. Проблемы синхронической и диахронической фонологии русского языка. СПб: Филологический факультет СПбГУ, 2004.

Statistic distribution of ä, ää, ö, öö, y, yy words by Google

word	%front
miljonääri	100
interiööri	100
primääri	100
afääri	100
revolutionäärinen	100
antipyriini	100
kasööri	100
hydrosfääri	100
sutenööri	98,5
stereotyyppi	98,5
konduktööri	98
amatööri	98
jonglööri	98
aldehydi	95,5
molekyyli	92,5
vulgääri	89
elektrolyysi	86,7
asetyleeni	86
metafyysinen	84
asymmetrinen	83,5
manikyristi	83
an-hydridi	80,5
manikyryri	80
hieroglyfi	80
polymeeri	77,5
trotyyli	75
synonyymi	61

apokryfi	55
analyysi	52
marttyyri	48
psykoanalyysi	40,5

Statistic distribution of e, ee, i, ii words by Google

word	%front
partenogeneettinen	100
pumpernikkeli	92,5
heksametri	86
kilometri	76,5
timotei	69
heterogeeninen	55
arkkitehti	32
barometri	29,5
paleoliittinen	26
adjektiivi	25,5
karamelli	24
idiotismi	21,5
aprilli	21,5
positiivinen	21
historiikki	17
invaliditeetti	16,5
mannekiini	16,5
krusifiksi	16
appelsiini	14,5
dynamiitti	12,5
artikkeli	12,5
reumatismi	12,5
hypoteesi	9
revolveri	7,5

progressiivinen	7
puustelli	4
ekumeeninen	3
juveniili	3
alkemisti	2,5
beduiini	2,5
amuletti	2
ateisti	1
krysanteemi	1
transitiivi	1
opportunisti	>1
turisti	>1
partikkeli	>1
katrilli	>1
ateljee	>1
aritmeettinen	>1
bolshevikki	>1
syntaksi	0
symptomi	0
tyranni	0
följetongi	0
kysta	0
fakiiri	0
matrikkeli	0
utilismi	0
instruktiivi	0
paralleeli	0
humoristi	0
individualismi	0
elatiivi	0
fakulteetti	0