Miklós Törkenczy

The phonotactics of Hungarian verbs

It is a well-documented property of the lexicon of a natural language that it may have a stratified structure. Phonologically, stratification manifests itself in the fact that lexical items that belong to different lexical strata (sublexicons) may display (partially) different phonological regularities. Probably the best known examples are Japanese (Itô & Mester 1995) and English (Chomsky & Halle 1968). The stratum-specific phonological regularities may involve alternations (captured by phonological rules) or phonotactic patterns (captured by phonotactic constraints: SSCs, MSCs, or sequence constraints). These sublexicons exist synchronically — they are not simply etymologically identical/similar sets of words.[1] In some cases sublexicons/lexical strata may also be identified independently of the phonological regularities, specifically, phonological rules and/or phonotactic constraints may be associated with a specific word-class.[2] Stress-assignment in English, for instance, follows two patterns (the 'verb pattern' and the 'noun' pattern) rather than a single homogeneous one (cf. Chomsky and Halle 1968). The phonotactics of the Classical Arabic verb is different from that of the noun (cf. McCarthy 1981). A similar situation exists in Hungarian where the phonotactics of verbs is more restrictive that the phonotactics of non-verbs (cf. Trón & Rebrus 2000, 2001, Rebrus & Trón 2002 2005, Törkenczy 2000, 2001) and where word-class membership partially determines when Lowering applies, cf. Siptár & Törkenczy 2000, Törkenczy 2004).

In this paper I shall discuss the phonotactic constraints that apply to a well-identifiable stratum of the Hungarian lexicon, the sublexicon of verbs.[3] There is an interesting phonotactic difference between verb stems and non-verb stems. A

1 The items in a given sublexicon may or may not be etymologically related, as is the case with the sublexicons in Japanese (Yamato, Sino-Yapanese, Foreign and Mimetic, cf. Itô & Mester 1995). 'Foreign' words of various origin may constitute a sublexicon in a language (as in Japanese), but that does not mean that *all* the words of foreign origin in that language necessarily belong to the foreign stratum/sublexicon since some words of foreign origin are not identifiable phonologically (e.g. *tánc* 'dance' in Hungarian).

2 There may be other independently identifiable strata. Recently there have been claims that place-names form a phonotactic sublexicon in Hungarian, cf. Rebrus & Trón (2002).

3 Note that it is no longer possible to add monomorphemic verb stems to the Hungarian lexicon. All new verbs (loans or otherwise) that enter the lexicon must have a verbal suffix (typically *-ol/-el/-öl/-l* or *-oz/-ez/-öz/-z*: *sztrájk**ol*** 'go on strike', *download**ol*** [dɔunlo:dol] 'download$_V$', *esemes**ez**(ik)* 'use SMS text messaging on a mobile phone'). Thus, monomorphemic verb stems are a closed system.

The Even Yearbook 7 (2006), Department of English Linguistics, Eötvös Loránd University, Budapest
ISSN 1218 – 8808, http://seas3.elte.hu/delg/publications/even, © *2006, Miklós Törkenczy*

non-verb stem is not identifiable as a non-verb on the basis of the string of segments it consists of alone (by a native speaker), while a verb *may* be identified as a verb. This is a unidirectional relationship: some strings could not be monomorphemic verb stems, but there are strings that could equally be monomorphemic verb stems or non-verb stems: *nyom* 'trace$_N$', *nyom* 'push$_V$'; *part* 'shore$_N$', *tart* 'hold$_V$'; *domb* 'hill$_N$', but */domb/$_V$.

The primary cue to phonotactic 'verb-ness' is the stem-final coda.[4] We shall first consider monomorphemic free[5] verb stems.

A monomorphemic free verb stem can end in any single consonant except /t^y, t^s, č, ǰ, ž, x/.[6] Of these the lack of final /ǰ/ is not surprising since there are hardly any items in the lexicon with final short /ǰ/: *bridge* /briǰː/, ?/briǰ/ 'bridge (card game)'. Disregarding /ǰ/, non-verbs/nouns can end in any single consonant, so the other gaps are specific to verb stems. Some generalisations could be made but it is unclear whether the gaps in the case of verbs are accidental or not. For instance, we could claim that there is a constraint banning affricates at the end of a monomorphemic verb stem, but then it seems that long affricates (as opposed to short ones) *are* possible in the same position: *metsz* [mɛt^{s}ː] 'cut', *edz* [ɛd^{z}ː] 'train'.[7] However, there are only three monomorphemic verb stems that end in a long affricate: *metsz*, *edz* and *pedz* 'begin to understand'. Since the geminates do exist, I take a 'soft' approach and consider that affricates are possible at the end of a verb stem (i.e. the gaps are accidental). Similarly, I take the lack of /t^y, ž/ to be accidental because their voiced/voiceless counterparts /d^y, š/, respectively, exist as non-branching codas in verbs (e.g. *hagy* 'leave', *ás* 'dig'). The lack of /x/

4 It is often claimed that monomorphemic verb stems cannot end in a vowel in Hungarian (e.g. Vago 1980, 1989, Kiefer 1994). This is based on the fact that *all* monomorphemic verb stems that end in a vowel in isolation are 'v-adding' stems, i.e. they end in /v/ before a vowel-initial suffix: *lő* 'shoot' ~ '*lö**v**-ök* 'shoot' (1sg pres. indef.) (as opposed to vowel-final nominal stems, which may be 'v-adding' stems or 'non-v-adding' major stems: *ló* 'horse' ~ *lo**v**-ak* 'horses', *mel**ó*** 'job' ~ *mel**ó**-k* 'jobs'). If we assume that a stem-final /v/ underlies the [v] ~∅ alternation (as is usual in the relevant analyses), then it follows that, underlyingly, verb stems can only end in a consonant. In the present analysis, however, we consider the [v] ~∅ alternation non-phonological (suppletive), cf. Siptár & Törkenczy (2000). Consequently, it does not hold that all verb stems end in a consonant since both the /v/-final and the vowel-final alternants of the relevant verb stems are present in the lexicon.

5 This excludes *-ik*-verbs and defective verbs.

6 Examples: /p/ *kap* 'get', /t/ *hat* 'inflence', /k/ *rak* 'load', /b/ *dob* 'throw', /d/ *ad* 'give', 'd^y' '*hagy*' 'leave', /g/ *rág* 'chew', /f/ *döf* 'thrust', /s/ *vesz* 'take', /š/ *mos* 'wash', /v/ *szív* 'suck', /z/ *néz* 'look', /m/ *nyom* 'push', /n/ *ken* 'smear', /n^y/ *hány* 'vomit', /l/ *él* 'live', /r/ *szűr* 'filter', /j/ *fúj* 'blow'.

7 Note that the affricate [d^{z}ː] is usually analysed as the surface reflex of the underlying cluster /dz/, cf. Siptár & Törkenczy (2000), Siptár (2002).

is specific to the verb and it is excluded as a geminate coda as well.[8] This is due to a constraint that excludes /x/ from the coda in the verb sublexicon:

(1) *(verb sublexicon)*
/x/ is not licensed in the coda.

Apart from the effect of (1) single codas are not radically more restricted in monomorphemic free verb stems than in non-verbs.

The situation is strikingly different if we examine verb stems that end in a consonant cluster. I have argued elsewhere (Törkenczy 1994, Törkenczy & Siptár 1999ab, Siptár & Törkenczy 2000) that monomorphemic final three-term consonant clusters are exceptional/irregular. There are very few monomorphemic words that do contain these clusters (the complete list is: /mps/ *mumpsz* 'mumps'; /nks/ /*szfinx* 'sphinx', *szkunksz* 'skunk'; /nst/ *dunszt* 'steam', *kunszt* 'trick'; /rst/ *karszt* 'karst', *verszt* 'verst'; /kst/ *szext* 'sixte'; /ršt/ *vurst* 'sausage'; /ršč/ *borscs* <Russian soup>; /jst/ *lejszt* 'hard work'; /jšt/ *mihelyst* 'as soon as'). It must be noted that *none* of these exceptional items are verbs.

(2) below shows the complete list of two-term clusters (including geminates) that monomorphemic free verb stems can end in together with the number of stems in our database[9] (in angled brackets) that contain each type of final cluster:

(2)

dː	<1>	nz	<1>
dᶻː	<2>	ŋg	<51>
gː	<3>	(nl	<1>)
tˢː	<1>	lt	<14>
st	<117>	ld	<7>
št	<1>	(lg	<1>)
zd	<2>	lː	<34>
žd	<1>	rt	<8>
mt	<1>	rd	<2>
(mz	<1>)	rː	<2>
nt	<100>	jt	<19>
nd	<4>		

8 Geminate /žː/ is also unattested in the coda generally, not just in verbs, geminate coda /xː/ occurs only in non-verbs, e.g. *pech* /pexː/ 'bad luck'.

9 For the computerised database used see Kornai (1986) and Siptár & Törkenczy (2000)

Examples:[10] *fedd* 'scold', *edz* 'train', *függ* 'hang', *metsz* 'cut', *oszt* 'distribute', *fest* 'paint', *kezd* 'begin', *esd* 'beg', *teremt* 'create', (*nemz* 'beget'), *önt* 'pour', *mond* 'say', vonz 'attract', *zeng* 'resound', (*ajánl* 'recommend'), *olt* 'put out (fire)', *küld* 'send', (*uralg* 'rule'), *vall* 'confess', *tart* 'hold', *hord* 'wear', *varr* 'sew', *hajt* 'drive'.

The parenthesized clusters in (2) above are only apparent (and are only included because they are cited in the literature). *Nemz* and *uralg* (the only stems with final /mz/ and /lg/) are obsolete in Educated Colloquial Hungarian as free stems and *ajánl* normally does not contain a final cluster [ɔjaːl(ː)]. I consider *teremt* is irregular/exceptional (and it may not even be monomorphemic (?)).

If we disregard these items, it becomes obvious that monomorphemic branching codas in the verb sublexicon are subject to much stricter constraints than (monomorphemic) branching codas in general. Here a monomorphemic branching coda must be a cluster that respects Sonority Sequencing (its right-hand term must be less sonorous than its neighbour on the left)[11] AND/OR is homorganic (it is a partial or a full geminate).

Specifically, such a verb-final cluster may be one of the following kinds: (i) it respects Sonority Sequencing, but is not strictly homorganic (/št, žd, jt/), or (ii) it respects Sonority Sequencing and is strictly homorganic (/st, zd, nt, nd, nz, ŋg, rt, rd, lt, ld/), or (iii) it is unlicensed by Sonority Sequencing and is strictly homorganic (/dː, tˢː, dᶻː, gː, lː, rː/). The generalisation that can be made is that clusters in set (i) are coronal while clusters in sets (ii) and (iii) are non-labial. This suggests that a branching coda in a monomorphemic free verb stem must be a coronal cluster if it is *only* licensed by government, but may be coronal or dorsal if it is licensed by place-binding or root-binding.[12]

Assuming the general condition on the licensing of Hungarian coda clusters according to which a branching coda must be licensed *either* by government *or* root-binding (cf. Siptár & Törkenczy 2000) and the universal right-to-left direction of government in the coda, this is captured by the following constraints:

10 For a complete list of verb stems representing the less populous types see Appendix I.

11 This is true of *regular* non-verb stems too (cf. Siptár & Törkenczy 2000), but in the verb sublexicon it is also *exceptionless*.

12 Throughout this paper I assume the system of constraints and phonotactic analysis proposed in Siptár & Törkenczy (2000). There *government* is defined as an asymmetrical relationship between segments where a segment X governs an adjacent segment Y if X is less sonorous than Y; and *binding* (which can apply between various nodes (e.g. between root-nodes or place nodes)) within the feature tree of a segments) as an asymmetrical relationship between segments where a bound segment contains dependent structure. The coda in Hungarian is maximally binary. It may be followed by an Appendix at the end of an analytic domain under special conditions (never in a monomorphemic verb).

(3) *Branching Coda Constraint (verb sublexicon)*

a. In a coda cluster C1C2 government can only apply iff C1 and C2 are both COR.
b. In a coda cluster C1C2 (root or place) binding can only apply if C1 and C2 are both COR or DOR

(3) does not permit branching codas like /lk, rv/ (non-verb examples: *halk* 'silent', *érv* 'argument') for verbs because they are unlicensed by government (sonority-wise they are ok, but they are non-coronal) and they are not licensed by binding (their terms do not share a root or a place node). Similarly, /pː, mb/ (non-verb examples: *csepp* 'drop', *domb* 'small hill') — although they do have a shared root note and a shared place node, respectively — are not well formed branching codas in a verb because they are not licensed by government according to (3a) and they are also not licensed by binding since they are LAB (3b). [nt, ŋg, jt] are well-formed branching codas in a free verb: [nt] is licensed by government *and* binding (sonority sequencing is ok, it is COR and homorganic: e.g. *ment* 'save'); [ŋg] is licensed by binding *only* (sonority sequencing is ok, but it is not COR, however, it is homorganic: e.g. *reng* 'quake'); [jt] is licensed by government *only* (it is non-homorganic/non-geminate but sonority sequencing is ok and it is COR: e.g. *hajt* 'drive')

Not all the branching codas permitted by (3) are well-formed. Some further restrictions apply. It can be seen in (2) that the second position can only be filled by a sonorant if it shares its root-node with the first (i.e. if it root-binds the first (=it is a geminate)): clusters like /jl, rn, rn^{y}, ln/ (non-verb examples: *fájl* 'file', *konszern* 'concern', *szörny* 'monster') are not well-formed branching codas in this stratum although they are licensed according to (3) by government and/or place-binding. This can be stated as:

(4) *(verb sublexicon)*

In a coda cluster C1C2 if C2=[+son], then C2 must root-bind C1.

Affricates only seem to be possible as geminates in the branching coda. They cannot occur in the first position in a branching coda when not root-bound, or in the second position when not root-binding the first: clusters like /t^{s}k, rt^{s}, nt^{s}/ (non-verb examples: *barack* 'peach', *perc* 'minute', *tánc* 'dance') are not well-formed branching codas. The lack of non-root-bound affricates (in the first position) is expected (and is not specific to the verb stratum)[13] since they are not licensed by government in this position by the general coda constraint that enforces sonority

13 What *is* special in this stratum is that there are no exceptions to the general constraint. Words like *barack* 'peach' are regarded exceptional by Siptár & Törkenczy (2000).

sequencing when binding does not apply (see footnote 12 above). The lack of non-root-binding affricates (in the second position) requires an additional constraint because clusters like /rtˢ, ntˢ/ are licensed according to (3) by government and/or place-binding. This can be stated as:

(5) *(verb sublexicon)*

In a coda cluster C1C2 if C2 is an affricate (i.e. C2 branches under its root node), then C2 must root-bind C1.

The occurrence of palatals is extremely restricted in the branching coda of a free verb stem: only [+continuant] palatals (/š, ž, j/) can occur and only when governed (i.e. in C1 position). [-continuant] palatals (/tʸ, dʸ, nʸ/) do not occur at all and [+continuant] palatals cannot occur in C2 position or as geminates. Codas like /nʸtʸ, rdʸ, rnʸ, rš, jž, tʸː, dʸː, nʸː, šː, jː/ (non-verb examples: *ponty* 'carp', *tárgy* 'object', *szörny* 'monster', *vers* 'poem', *pajzs* 'shield', *pötty* 'polka dot', *meggy* 'sour cherry', *genny* 'pus', *friss* 'fresh', *gally* 'branch) are not permitted. This can be stated as:[14]

(6) *(verb sublexicon)*

In a coda cluster a [COR, -ant] segment must be [+cont] and governed.

There are some branching codas that do not occur, but are permitted by (1), (3), (4), (5), (6). We consider these accidental gaps. Some of the accidentally missing coda clusters, although not found in monomorphemic free verb roots, actually occur *undivided by a morpheme boundary* finally in free suffixed verbs: e.g. [ŋk] *kapar-u**nk*** 'scratch' (1pl pres indef.), [tː] *vezet-e**tt*** 'drive' (3sg past indef.). By contrast, no *morphologically undivided* coda cluster occurs verb-finally that should violate the constraints discussed above (e.g. *-bb* [bː] could not be a verbal suffix; non-verb-example: comparative *-bb nagy-o**bb*** 'bigger'). This supports the accidental gap interpretation of the unattested clusters that are permitted by (1), (3), (4), (5) (6).

We have seen that the constraints that apply to branching codas in the verb sublexicon are more restrictive than the general constraints on branching codas.

14 Note that [-continuant] geminate palatals (/tʸː, dʸː, nʸː/) do occur finally when they are *morphologically complex.* This happens when the imperative suffix *-j* is attached to a verb stem that ends in /d, dʸ, n, nʸ, l, j/ and palatalisation applies (or simple concatenation in the case of /j/): ***ad-j*** [ɔd̲ʸː] 'give!', ***hagy-j*** [hɔd̲ʸː] 'permit!', ***ken-j*** [kɛn̲ʸː] 'smear!', ***hány-j*** [haːn̲ʸː] 'vomit!', ***él-j*** [eːj̲ː] 'leave', ***fúj-j*** [fuːj̲ː] 'blow!'(cf. Siptár & Törkenczy 2000). The constraints that we have discussed apply to monomorphemic verbs.

The same effect can be seen in the case of intervocalic clusters (albeit in a weaker form).

There are no monomorphemic verbs with an intervocalic consonant cluster consisting of more than three consonants. There are a few examples in our database of verb stems that seem to contain an intervocalic three-term cluster that is undivided by a morpheme boundary. An exhaustive list is given in (7):

(7)[15]

ksp	*exponál*
str	*kasztrál*, *nosztrifikál*, *regisztrál*
štr	*kimustrál, ministrál*
mpr	*impregnál, imprimál, improvizál*
mdl	*vámdliz*
nsp	*transzponál*
nšp	*inspiciál*
nt^{s}v	*kifiruncvancigol* (?)
ŋkr	*inkriminál*
rtr	*pertraktál*

Note that in (7) *all* 'monomorphemic' verbs end in -*Vl* (or in one case), -*Vz*. These endings are (identical with) denominal verb-forming derivational suffixes that productively attach to clearly identifiable stems: e.g. *kasza* 'scythe$_N$'– *kaszá-**l*** 'scythe$_V$, *dob* 'drum$_N$' – *dob-**ol*** 'drum$_V$', *gumi* 'rubber$_N$' – *gumi-**z*** 'rubberize'. Even though the verbs in (7) do not have such an easily identifiable nominal root+verb-forming suffix structure (**expona – exponál* 'release the shutter of a camera'), it is reasonable to assume that they follow the same pattern: they consist of a nominal 'phantom stem' and a *denominal* verb-forming derivational suffix.[16] Under this analysis there are no examples of monomorphemic verbs with three-member intervocalic consonant clusters. Following Törkenczy & Siptár 1999ab, Siptár & Törkenczy 2000 (see the arguments therein) I consider monomorphemic intervocalic CCC clusters phonotactically ill-formed in general, and thus *all* monomorphemic lexical items with an intervocalic cluster of more than two members are exceptional/irregular. Thus, the restriction is not specific to verbs,

15 Glosses: *exponál* 'release the shutter of a camera', *kasztrál* 'castrate', *nosztrifikál* 'validate a foreign diploma', *regisztrál* 'register$_V$', *kimustrál* 'discard', *ministrál* 'assist a priest at mass', *impregnál* 'impregnate', *imprimál* 'pass the proofs for the press', *improvizál* 'improvise', *transzponál* 'transpose', *inspiciál* 'inspect', *inkriminál* 'incriminate', *pertraktál* 'discuss in detail'. *Kifiruncvancigol* 'figure out, calculate' and *vámdliz* '?' are only included because they can be found in the database. I have been unable to establish if *firuncvancig* and *vámdli* exist as nouns (they are not in the database) and what *vámdliz* means.

16 They are like the 'phantom stems' in *in**sist**, con**sist**, re**sist***, etc. in English, cf. Chomsky & Halle (1968), Aronoff (1976).

but again, verb stems do not even offer irregular counterexamples to the generalisation.

Let us now examine two-member intervocalic clusters in verb stems. Table I below shows the intervocalic CC clusters in monomorphemic verbs attested in the database and compares them to the attested CC clusters in monomorphemic non-verbs. The notation used is the following: a blank space in an intersection of a row and a column means that the relevant cluster is unattested or only occurs when the two consonants are separated by an analytic morphological domain boundary; a '+' occurs if it is attested in monomorphemic items and the number of such items in the database is n>15; numbers have been used to indicate the number of monomorphemic items in the database when the cluster in question is attested in monomorphemic items and the number of such items is n≤15. *Colons* separate data about non-verbs from data about verbs: non-verb data appear on the left of the colon (*non-verb:verb*). The symbol '−' is used to indicate the lack of a cluster if the same cluster is attested on the other side of the colon ('−*:n*', or '*n:* −'), i.e. if the cluster is attested in a monomorphemic verb *or* non-verb only. For ease of reference intervocalic clusters that occur in monomorphemic verb stems (i.e. those with a number or a '+' on the right of the colon: '*:n*' or '*:*+') have been encircled (in any colour). The circle is **red** if the number of such items in the database is n>15 (i.e. if a '+' appears on the right of the colon: '**:+**'); and the circle is **black** if the number of such items in the database is n≤15 (i.e. if a number appears on the right of the colon: '*:n*') *while* the number of non-verb items is greater than the number of verb items (i.e. '+*:n*' or '*m:n*' *where m>n*). The circle is **blue** if the number of verb items is greater than the number of non-verb items and the number of non-verb items is zero (i.e. '−*:n*'); and the circle is **green** if the number of verb items is greater than the number of non-verb items and non-verb items do exist (i.e. '*m:n*' *where m<n and m≠0*).

Some explanation is in order about the data included in Table I before we examine it. I have included all verb stems in the database that can be regarded *suspicious* of having an internal monomorphemic intervocalic CC cluster. I have relied on the following guidelines to judge what is suspicious:

(8) In order to qualify as a verb-stem-internal cluster ('vsi-cluster'), an intervocalic cluster must meet one or more of the following partially overlapping, partially conflicting criteria:

(a) *The position of the intervocalic cluster*
The VCCV string must be contained in a verb stem: the CC cluster cannot be divided or immediately preceded/followed by the edge of a *productive* suffix (and the stem must have more-or-less non-transparent, non-compositional meaning).
Example: this rules out /kt/ in *bu**kt**at* 'flunk' (because *-tat* is a productive suffix whose edge breaks up the cluster: *bu**k-t**at*), and /kl/ in *csu**kl**ik* 'hiccup'(because *-ik* is a productive suffix whose edge immediately follows the cluster: *csu**kl-**ik*).

(b) *The formal properties of the internal stem*
The verb-stem that contains the CC cluster may or may not have an internal stem, but if it does, then the pre-ending (ghost) stem (i) may only have bound forms (in the given meaning) or (ii) must be suppletive if one of its allomorphs is free.
Example: this rules out /pr/ in *a**pr**-ít* 'chop up' (because the stem has a free allomorph *apró* 'small'), /mp, st/ in *ko**mp**osz**t**-ál* 'compost$_V$'(because *komposzt* 'compost$_N$' is a free morpheme), and /rm/ in *ka**rm**-ol* 'claw$_V$' (because the stem has a free allomorph *karom* 'claw$_N$').

(c) *The interpretation of the verb stem containing the intervocalic cluster*
The CC cluster must be truly morpheme-internal (wholly contained) within a strict or extended *verb stem*.
Example: this rules out /ng/ in *angoloz* 'study English' (because, although the cluster is wholly contained within a monomorphemic stem, it is not a verb stem: [*a**ng**ol*]$_N$ *oz*]$_V$)

(i) *verb stem = strict monomorphemic verb root*
The VCCV string must be strictly verb-root-internal: the CC cluster may not be preceded/divided/followed by the edge of even an unproductive suffix/ending.
Example: this accepts /lv/ in *o**lv**as* 'read' (because it is a monomorphemic root and does not contain even a ghost stem), but rules out /pt/ in *ka**pt**at* 'climb with difficulty' and /ks/ in *ma**x**imál* 'maximize' (because they are extended verb stems, i.e. they contain ghost stems <u>*kap*</u>*-tat*[17] <u>*maxim*</u>*-ál* or <u>*maximá*</u>*-l*)

17 I take the unproductive ending *-tat* here to be different from the productive causative *-tat/-tet* suffix.

(ii) *verb stem = extended verb root (the ghost stem must be verbal)*
The CC cluster may be part of a ghost stem, but cannot be part of a non-verbal ghost stem: the cluster may be preceded/divided/followed by the edge of an unproductive suffix/ending iff the unproductive verb-forming suffix/ending is deverbal (verb→verb).
Example: this accepts /lv/ in *o**lv**as* (because it is a monomorphemic root and does not contain even a ghost stem) and /pt/ in *ka**pt**at* (because the ghost stem in *kap*-*tat* is verbal, i.e. the unproductive verb-forming ending *-tat* is a deverbal one), but rules out /ks/ in *ma**x**imál* (because the ghost stem in *maxim*-*ál* or *maximá*-*l* is non-verbal since the unproductive verb-forming ending *-(V)l* is not a deverbal one)

(iii) *verb stem = extended verb root (the ghost stem may be non-verbal)*
The CC cluster may be part of a non-verbal ghost stem: The cluster may be preceded/divided/followed by the edge of an unproductive verb-forming suffix/ending (X→verb).
Example: this accepts /lv/ in *o**lv**as* 'read', /pt/ in *ka**pt**at* 'climb with difficulty' and /ks/ in *ma**x**imál*

I have included *all* suspicious/potential intervocalic vsi-clusters in Table I. All verb stems containing a suspicious/potential intervocalic vsi-cluster must meet the criteria (8a, b) in order to qualify. Those that do have been sorted into three sets: 'strictly suspicious' ('strict'), 'moderately suspicious' ('moderate') and 'loosely suspicious' ('loose'). A *strict intervocalic vsi-cluster* is one that can be found in at least one verb that meets the criteria (8a, b, ci) (e.g. /lv/ *olvas*). A *moderate intervocalic vsi-cluster* is one that can be found in at least one verb that meets the criteria (8a, b, cii) (e.g. /pt/ *kaptat*). A *loose intervocalic vsi-cluster* is one that can be found in at least one verb that meets the criteria (8a, b, ciii) (e.g. /ks/ *maximál*).[18] In Table I, the markings for *strict and moderate* intervocalic vsi-clusters have been highlighted by yellow shading.

The distinction between the moderate and the loose set is an extremely difficult (if not impossible) one to make (consistently) since only non-productive verb-forming endings added to ghost stems qualify for being sorted into deverbal and non-deverbal. I took an (arbitrary, authoritarian) and procedural approach to this problem. I have sorted the suspicious verb-forming endings into three sets. I have considered an unproductive verb-forming ending consistently *deverbal* if it is listed as such in *all* the four (very different) sources I have used (Kiefer 1998,

18 These sets are in a relationship of inclusion: a strictly suspicious cluster is also moderately suspicious and loosely suspicious, etc. Since it makes no difference in the discussion we shall refer to the strict, moderate and loose sets as if they were complementary.

Rebrus 2000*a,* Tompa 1961, Velcsovné 1988). Similarly, I have considered an unproductive ending consistently *non-deverbal* if it is listed as such in *all* the four sources. Finally, I have made no decision and judged an ending *uncertain* if it was not possible to make a decision. The latter state of affairs often occurs when the stem is bound: *-kVdik* in [*tanár*]$_{N}$ *kodik*]$_{V}$ 'work as a teacher' is obviously non-deverbal since *tanár* 'teacher' is a noun, it is deverbal in [*emel*]$_{V}$ *kedik*]$_{V}$ 'rise' since *emel* 'lift' is a verb, but in [*incsel*]$_{?}$ *kedik*]$_{V}$ 'tease' it is uncertain since *incsel-* does not occur elsewhere/is bound. Similar examples: *-ít*: [*meleg*]$_{ADJ}$ *ít*]$_{V}$ 'warm up'(*meleg* 'hot'), [*áll*]$_{V}$ *ít*]$_{V}$ 'make stop' (*áll* 'stop$_{V}$'), but [*sand*]$_{?}$ *ít*]$_{V}$ 'squint', [*közvet*]$_{?}$ *ít*]$_{V}$ 'mediate'; *-ul/-ül*: [*tan*]$_{N}$ *ul*]$_{V}$ 'study' (*tan* 'teaching$_{N}$'), [*nyom*]$_{V}$ *ul*]$_{V}$ 'push ahead' (*nyom* 'push'), but [*izg*]$_{?}$ *ul*]$_{V}$ 'be anxious'. Thus, *-kodik/-kedik/-ködik*, *-ít*, and *-ul* are judged uncertain. I have considered an item to contain a cluster that is a member of the *moderate* set if its ending was consistently deverbal, I have judged it *loose* if its ending was consistently non-deverbal *or* uncertain.

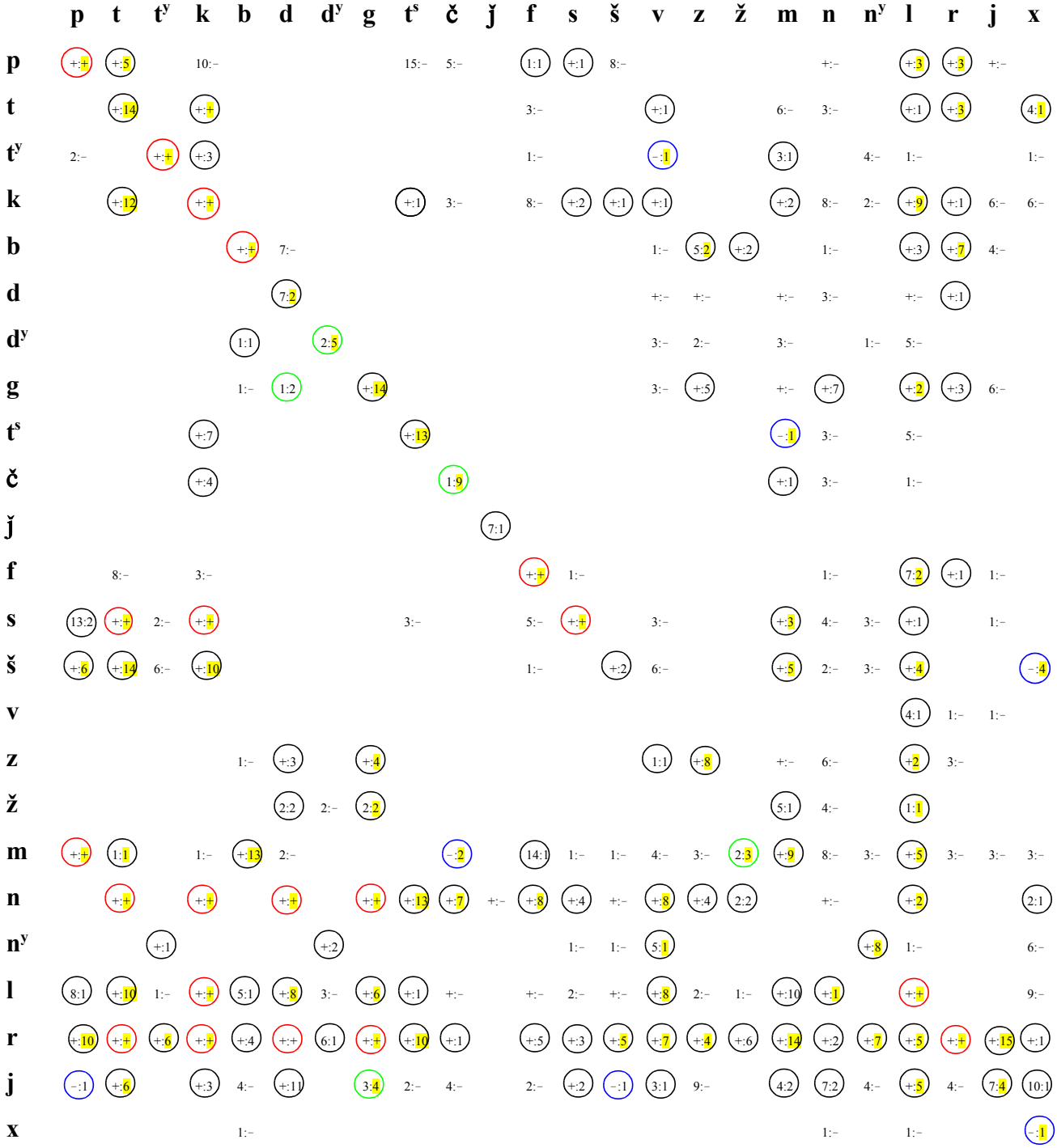
The list of the relevant non-productive endings sorted is the following (only those variants have been included that have been found in the database in stems with a suspicious vsi-cluster):

(9) *Endings*

deverbal	*-(k)ódik, −(k)ózik, −(k)őzik, −(t)at/−(t)et, −ad/−ed, −an/−en, −ant/-ent, −aszt/−eszt, −dogál/-degél, −dos, −ell/−all, −gat/−get, −int, −lal/−lel, −ó(d)zik, −ódik/-ődik, −og/−eg/−ög, -ong/−eng/−öng*;
non-deverbal	*−ál/−él/−l, −ol/−el/−öl/−al/−l, -oz/-ez/-öz/-az/−z*;
uncertain	*−(k)odik/−(k)odik/−(o)dik, −(k)ozik/−(k)ezik/−(k)özik, -ász(ik)/−ész(ik), −ít, −ul/−ül*.

The examples that follow Table I are always the 'best' ones in the sense that a strict example has been given for a cluster is there is one, a moderate example has been given if there is not a strict example, but a moderate one exists and finally, a loose example is only provided if there are no strict or moderate examples. **Strict** examples have been set in bold, <u>moderate</u> examples have been underlined and endings have been CAPITALISED.

Table I *Intervocalic CC clusters in monomorphemic non-verbs and verbs*[19]



19 For an exhaustive list of verbs containing these clusters, see Appendix II.

Examples: ***lappang*** 'incubate', *kapTAT* 'climb with difficulty', *klopfOL* 'tenderize meat by beating', *abszolvÁL* 'complete', *átcaplAT* 'pop over to', *töprENG* 'speculate'; *suttOG* 'whisper', *vetKŐZIK* 'undress', *ötvÖZ* 'alloy', ?*csatlAKOZIK* 'join', *kotrÓDIK* 'shove off', *rothAD* 'rot'; *kettyINT* 'copulate', *bütykÖL* 'fix', *kotyvASZT* 'concoct', *fitymÁL* 'belittle'; *bakTAT* 'trudge', *csökkEN* 'reduce', *akceptÁL* 'appreciate', *fixírOZ* 'eye', *kuksOL* 'crouch', *likvidÁL* 'liquidate', *lakmárOZIK* 'feast on', *zaklAT* 'harass', *akkreditÁL* 'accredit'; *meghibbAN* 'go mad', *tobzÓDIK* 'luxuriate', *lebzsEL* 'idle', *öblÍT* 'rinse', *ébrED* 'wake up'; *idDOGÁL* 'drink a little', *addresszÁL* 'address'; *buggyAN* 'go bad'; *szökdécsEL* 'jump about', *nyaggAT* 'nag', *vegzÁL* 'annoy', *regnÁL* 'reign', *fogLAL* 'occupy', *emigrÁL* 'emigrate'; *lubickOL* 'splash about', *moccAN* 'move', ***kecmerEG*** 'fumble'; *pocskondiáZ* 'disparage', *becsiccsENT* 'become tipsy', *becsmérEL* 'slander'; menedzsEL 'manage'; *puffAD* 'become bloated', *caflAT* 'shamble'; *diszponÁL* 'dispose of sg', *vesztEGET* 'bribe', *viszkET* 'itch', *bosszANT* 'annoy', piszmOG 'fiddle with sg', *észlEL* 'percieve'; *tespED* 'languish', ***sisterEG*** 'sizzle', *áskálÓDIK* 'scheme', *eltussOL* 'cover up', ***ismer*** 'know', *koslAT* 'court', *poshAD* 'become stagnant'; *mozdÍT* 'move', *igazGAT* 'direct', *közvetÍT* 'mediate', *duzzOG* 'sulk', ?*ízLEL* 'taste'; *pezsdÍT* 'liven up', *rebesGET* 'rumour', *prézsmitÁL* 'preach', ?*vizslAT* 'examine'; ***hömpölyÖG*** 'surge', ?*nyomtat* 'print', ***csámborOG*** 'wander aimlessly', *csámcsOG* 'eat noisily', *somfordÁL* 'slink', *hemzsEG* 'swarm', *brummOG* 'growl', *emlEGET* 'mention repeatedly'; **henterEG** 'welter', lankAD '', ***sündörÖG*** 'slink', *engED* 'allow', *hencEG* 'boast', ***kuncsorOG*** 'beg', ***ténferEG*** 'hang around', *vonszOL* 'drag', *szenvED* 'suffer', *inzultÁL* 'insult', *avanzsÁL* 'rise in rank', ?*kínlÓDIK* 'suffer', *inhalÁL* 'breathe in medicine'; *szontyolODIK* 'become sad', *göngyÖL* 'wrap', *senyvED* 'suffer', *sunnyOG* 'slink'; *kolportÁL* 'spread rumours', *álTAT* 'delude', ***alkot*** 'create', *kiakolbódÍT* 'kick sb out', ***oldalOG*** 'slink', *találGAT* 'guess', ?*lefalcOL* 'escape', ***olvas*** 'read', *alkalmAZ* 'employ', ?***volna*** <function word expressing unreal condition>, ***ballag*** 'amble'; *torpAN* 'stop suddenly', ***történik*** 'happen', *hortyOG* 'snore', *serkENT* 'stimulate', *dorbézOL* 'carouse', *kérdEZ* 'ask', *meggárgyUL* 'go soft in the head', *kergET* 'chase', *sercEG* 'sizzle', *szürcsÖL* 'drink noisily', *morfondírOZ* 'ponder', *forszírOZ* 'insist', ***társalOG*** 'discuss', ?***örvend*** 'be glad', *borzONG* 'shiver', *horzsOL* 'scratch', *förmED* 'snarl', *internÁL* 'relocate by force', *görnyED* 'bend', *birLAL* 'hold <legal>', *virrASZT* 'stay awake', *gerjED* 'become aroused', *archaizÁL* 'make sg archaic'; *selypÍT* 'lisp', ***kujtorOG*** 'loiter', *hajkurÁSZIK* 'chase after sg', *bolydUL* 'stir', *bolyGAT* 'disturb', *majszOL* 'munch', *hajsókÁL* 'swing', ?*jajveszékEL* 'wail', ?*lejmOL* 'bum', *ajnárOZ* 'adore', *fejlESZT* 'develop', vijjOG 'screech', ?*megújhODIK* 'be reborn'; kehhENT 'cough';

It can be seen in Table I that, although the set of intervocalic CC clusters that can be found in a monomorphemic verb stem (strict, moderate or loose) is similar to the set of intervocalic CC clusters in non-verb stems, verb stems are phonotactically more restrictive than non-verb stem in a 'weak' (frequency-related) sense. The intervocalic CC clusters found in verbs do follow the general pattern for interconstituent clusters (i.e. they conform to the general interconstituent constraints, compare Siptár & Törkenczy 2000) and no additional, more restrictive phonotactic constraints can be formulated with a sufficient degree of generality in terms of natural classes. The higher restrictiveness of verb phonotactics manifests itself in the 'population' (type frequency) of clusters within the phonotactic space permitted by the general constraints.[20]

The following generalisations can be made: (a) the number of the types of clusters that can be found in verbs is significantly lower than those found in non-verbs. Of the total 249 attested monomorphemic intervocalic CC clusters, 242 are attested in non-verbs and only 145 are attested in verbs; (b) It is extremely rare for an attested intervocalic CC cluster to be found in more verb stems than non-verb stems, i.e. *nv*<*v* (where *nv* is 'non-verb' and *v* is 'verb'). Of the total 249 attested monomorphemic intervocalic CC clusters, there are only 12 clusters of this kind, 7 of which are only attested in verbs (numbers in parentheses refer to the *nv*:*v* ratio): /t^{y}v (−:1), t^{s}m (−:1), sx (−:4), mč (−:2), jp (−:1), jš (−:1), xx (−:1)/, /d^{y}ː (2:5), gd (1:2), čː (1:9), mž (2:3), jg (3:4)/, and the difference in favour of verbs is typically very small (8 times out of the 12 cases *v*−*nv*=1); (c) By contrast, it is very often the case that an attested cluster is found in more non-verb stems than verb stems (or that it occurs in non-verb stems only), i.e. *nv*>*v*. Of the 249 attested intervocalic CC clusters 209 occur in a higher number of non-verb stems than verb stems, of which 104 are only attested in non-verbs. Also the difference in favour of non-verbs is typically not small: in 92 cases out of the 209 attested clusters the difference in favour of non-verb stems is at least 10 (*nv*−*v*≥10, where non-occurrence is counted as zero); (d) There are a significantly greater number of 'populous' cluster types (i.e. a 'populous' cluster type is a

20 Note that frequency (type or token) has no (clear) formal/theoretical status in a generative model of phonology (derivational generative phonology, Optimality Theory or Government/CV/VC Phonology) — even though it is sometimes referred to in various generative analyses. Thus, inasmuch as the observations made here are correct and relevant to phonotactics, they actually point/lead outside the model used in this dissertation. Frequency *is* accommodated by other frameworks, e.g. in pre-generative (structuralist) phonology (e.g. Spang-Hanssen 1959) and recently, in laboratory phonology (e.g. Hay, Pierrehumbert and Beckman 2003, Pierrehumbert 1994), the (re)emerging functional approach (Bybee 2001, Bybee & Hopper 2001, Rebrus & Trón 2002, Trón & Rebrus 2000, 2001) and elsewhere (e.g. Hayes & Cziráky Londe 2005).

cluster found in a relatively high number of stems) in the case of non-verbs than in the case of verbs. Of the 249 attested intervocalic CC clusters there are 124 relatively populous non-verb clusters (clusters with at least 10 tokens: *nv:* where *nv*≥10) and only 36 relatively populous verb clusters (*:v* where *v*≥10). Interestingly, the distribution of the *populous* intervocalic CC clusters in the phonotactic space is somewhat different for verbs and non-verbs. In the case of non-verbs the populous clusters are distributed more or less evenly in the phonotactic space delimited by the relevant constraints (cf. Siptár & Törkenczy 2000). By contrast, in the case of verb stems, populous clusters tend to be geminates, s/š+stop clusters[21] and sonorant+obstruent clusters, but not obstruent+sonorant clusters (which can be populous in non-verbs). This may suggest that, although the Syllable Contact Law does not hold in Hungarian in general, the phonotactics of verbs is skewed towards it because the populous intervocalic clusters tend to conform to it.[22]

These effects become even stronger the stricter one is in circumscribing what qualifies as an intervocalic cluster in a monomorphemic verb: of the 249 attested intervocalic CC clusters in Table I 145 are attested in verbs under the loose interpretation.[23] Of these only 84 are strict or moderate vsi-clusters (highlighted by yellow shading in Table I[24]), of which only 19 are strict vsi-clusters. An exhaustive list of the strict vsi-clusters together with all the strictly suspicious stems is given in (10):

21 Stop+stop is also populous in verbs provided that /t/ and /k/ combine.

22 Note that this does not imply that there are no clusters in verbs that violate the Syllable Contact Law, just that those types of clusters are not populous in verbs.

23 Assuming that what is an intervocalic cluster in a monomorphemic verb under a strict or moderate interpretation is also an intervocalic cluster in a monomorphemic verb under a loose interpretation.

24 Note that the highlighted numbers in Table I are misleading to a degree because not *all* verb stems exemplifying a given type of moderate or strict cluster necessarily contain a moderate or strict vsi cluster. The number is highlighted if the cluster is attested in at least *one* moderately or strictly suspicious verb stem.

(10) a. t^{s}m ke**cm**erEG ‘fumble’
št si**st**erEG ‘sizzle’
šm i**sm**er ‘know’
mp he**mp**erEG ‘welter’, so**mp**olyOG ‘sneak’, hö**mp**ölyÖG ‘surge’
nt he**nt**erEG ‘welter’, tá**nt**orOG ‘stagger’
mb csá**mb**orOG ‘wander aimlessly’, i**mb**olyOG ‘waver’
nd sü**nd**örÖG ‘slink’
nč ko**ncs**orOG ‘beg’, ku**ncs**orOG ‘beg’
nf té**nf**erEG ‘hang around’
lk a**lk**ot ‘create’, a**lk**uszIK? ‘bargain’
ld o**ld**alOG? ‘slink’
lv o**lv**as ‘read’
ln vo**ln**a? <function word expressing unreal condition>
rt ta**rt**ózTAT? ‘delay’, tö**rt**énIK ‘happen’, zsö**rt**ölŐDIK ‘grumble’
rš tá**rs**alOG ‘discuss’
rv ö**rv**end? ‘be glad’
jt ku**jt**orOG ‘loiter’

b[25] pː la**pp**ang ‘incubate’
lː ba**ll**ag ‘amble’, vá[l(ː)]al

Thus, we have seen that the phonotactics of monomorphemic verb stems is more restrictive than the phonotactics of non-verb stems, which manifests itself (i) in the stricter constraints that apply to the coda and (ii) the frequency effects intervocalic consonant clusters display.

It would be desirable to relate the fact that stricter coda constraints apply in monomorphemic verbs to a difference between possible morphologically complex codas available for nouns and for verbs. If we examine word-final codas that consist of a stem-final consonant and a consonant that realises a synthetic suffix, we find that the codas of this structure that can occur at the end of verbs is a subset of those that can occur at the end of nouns. The suffixes in question are the past tense suffix in the case of verbs and the accusative in the case of nouns/adjectives. Both of them are realised as /t/ when they can form a branching coda with a stem-final consonant, otherwise they are preceded by a ‘linking’ vowel (see the details in e.g. Siptár & Törkenczy 2000, Törkenczy & Siptár 2000,

25 Many of the strict and moderate verb stems with an intervocalic geminate have an ‘onomatopoeic’ ‘expressive’ or ‘mimetic’ character (cf. Appendices I and II). I shall not pursue this idea here, but it is possible that these kinds of words have a special phonotactic makeup, perhaps they form another phonotactic stratum (cf. Fudge 1970, Itô & Mester 1995).

Törkenczy 2004). The coda clusters that can be formed in this way are given below for nouns and for verbs:

(11)

nouns (accusative)	*verbs* (past)
st	–
št	–
nt	nt
n^{y}t	n^{y}t
lt	lt
rt	rt
jt	jt

As can be seen in (11) the phonotactics of morphologically complex branching codas are very similar in nouns and verbs except that it is more restrictive in verbs, because in a verb the first consonant must be a sonorant: /st, št/ can only occur as codas of this kind in a noun: *csempész-t* 'smuggler' (acc.) vs. *csempész-ett*, **csempész-t* 'smuggle' (3sg past indef.); *les-t* 'raised hide' (acc.) vs. *les-ett*, **les-t* 'spy on sb.' (3sg past indef.). Notice, however, that although the relationship between non-verb and verb phonotactics is the same in morphologically simplex branching codas and morphologically complex branching codas (in both cases it is the verb phonotactics that is more restrictive), in this case it is not possible to derive the behaviour of the morphologically complex branching codas in verbs from that of the morphologically simplex ones since /st, št/ are well-formed final codas in monomorphemic verbs: /st/ *oszt* 'distribute', /št/ *fest* 'paint' (and /st/ is even the most populous type, cf (2) above). Thus, the restriction on morphologically complex branching codas in verbs must be an extra stipulation.[26]

26 Compare Rebrus & Trón (2005)

Appendices: lists

Note: exhaustive lists are marked [full]; the number of tokens (stems) in a given type appears in angled brackets: <>

I. Monomorphemic free CC-final verb stems

Note: some if the items containing the codas /st, nt, lː/ are/may be analysed as polymorphemic

dː	<1>	*fedd* [full]
d^{z}ː	<2>	*edz, pedz* [full]
gː	<3>	*csügg, függ* [full]
t^{s}ː	<1>	*metsz* [full]
st	<117>	*aggaszt, akaszt, alátámaszt, alvaszt, apaszt, áraszt, bágyaszt, biggyeszt, bomlaszt, borzaszt, ...*
št	<1>	*fest* [full]
zd	<2>	*kezd, küzd* [full]
žd	<1>	*esd* [full]
mt	<1>	*teremt* [full]
(mz	<1>)	*nemz* [full]
nt	<100>	*bánt, biccent, billent, bont, bosszant, bólint, böffent, buffant, buggyant, csahint,...*
nd	<4>	*csikland, fecskend, mond, örvend* [full]
nz	<1>	*vonz* [full]
ŋg	<51>	*állong, bolyong, borong, borzong, búsong, csapong, cselleng, cseng, csüng, dong, dühöng, dülöng, eseng, feszeng, fetreng, forrong, hajlong, hőzöng, hullong, ing, jajong, kering, kong-bong, lappang, lázong, leng, lézeng, mereng, ömleng, ődöng, őrjöng, rajong, ráng, ring, setteng, sikong, szállong, szorong, teng, teng-leng, tolong, töpreng, ujjong, verseng, villong, visong, zajong, zeng, zsibong, zsong, zsong-bong* [full]
(nl	<1>)	*ajánl* [full]
lt	<14>	*félt, kelt, kiált, költ, olt, ölt, rikolt, sikolt, süvölt, tilt, tölt, üvölt, vált, volt* [full]
ld	<7>	*áld, fold, küld, old, száguld, told, told-fold* [full]
(lg	<1>)	*uralg* [full]
lː	<34>	*áll, csekélyell, dall, drágáll, furcsáll, fuvall, gyengéll, hall, hátall, hosszall, hull, javall, kall, kell, kevesell, késől, koráll, lövell, nagyoll, nehezell, pall, restell, rivall, rosszall, rühell,*

		sarkall, sokall, sugall, száll, szégyell, szökell, vall, vénell, vétkell [full]
rt	<8>	*árt, ért, gyárt, irt, kísért, márt, sért, tart* [full]
rd	<2>	*hord, kérd* [full]
r:	<2>	*forr, varr* [full]
jt	<19>	*bujt, ejt, fejt, felejt, fojt, gyújt, gyűjt, hajt, lejt, nyújt, ojt, óhajt, rejt, sejt, sóhajt, sújt, szakajt, szalajt, veszejt* [full]

II. Verb stems containing intervocalic vsi-clusters

Note: **Strict** examples have been set in bold, moderate examples have been underlined and endings have been CAPITALISED.

(a) vsi-cluster=non-geminate <n=735>

pt	<5>	*csípTET, kapTAT, adoptÁL, akceptÁL, kooptÁL* [full]
pf	<2>	*cupfOL?, klopfOL* [full]
ps	<1>	*abszolvÁL* [full]
pl	<3>	*átcaplAT, táplÁL, kopLAL* [full]
pr	<3>	*deprimÁL, töprENG, reprodukÁL* [full]
tk		*vetKŐZIK, követKEZIK, kotkodácsOL, hagyatKOZIK ....*
tv	<1>	*ötvÖZ*, [full]
tl	<1>	*csatlAKOZIK*? [full]
tr	<3>	*kotrÓDIK, fetrENG, patronÁL* [full]
th	<1>	*rothAD* [full]
t^yk	<3>	*bütykÖL, lötykÖL, ügyKÖDIK* [full]
t^yv	<1>	*kotyvASZT* [full]
t^ym	<1>	*fitymÁL*, [full]
kt	<12>	*affektÁL, annektÁL, bakTAT, ikTAT, diktÁL, kecsegTET, lükTET, nyugTAT?, paktÁL, praktizÁL, rugTAT, vágTAT* [full]
kt^s	<1>	*akceptÁL* [full]
ks	<2>	*fixírOZ, maximÁL* [full]
kš	<1>	*kuksOL* [full]
kv	<1>	*likvidÁL* [full]
km	<2>	*lakmárOZIK, tukmÁL* [full]
kl	<9>	*akklimatizÁL, csakliZ, csiklandOZ, deklamÁL, érdeklŐDIK, öklEL, proklamÁL, reklamÁL, zaklAT* [full]
kr	<1>	*akkreditÁL* [full]
bz	<2>	*képzEL, tobzÓDIK* [full]

bž	<2>	*habzsOL, lebzsEL* [full]
bl	<3>	*öblÍT, téblábOL, szublimÁL* [full]
br	<7>	*babrÁL, celebrÁL, ebrudAL, ébrED, fabrikÁL, vibrÁL, zabrÁL* [full]
dr	<1>	*addresszÁL* [full]
dʸb	<1>	*lögybÖL* [full]
gd	<2>	*szökdEL, szökdécsEL* [full]
gz	<5>	*egzaminÁL, egzecírOZ, egzisztÁL, rögzÍT, vegzÁL* [full]
gn	<7>	*agnoszkÁL, dezignÁL, ignorÁL, impregnÁL, regnÁL, stagnÁL, szignÁL* [full]
gl	<2>	*fogLAL, tagLAL* [full]
gr	<3>	*degradÁL, emigrÁL, integrÁL* [full]
tˢk	<7>	*dömöckÖL, evickÉL, kéredzKEDIK, lubickOL, packáZ, peckEL, pöckÖL* [full]
tˢm	<1>	***kecmerEG*** [full]
čk	<4>	*fröcskÖL, pacskOL, pocskondiáZ, tapicskÁL* [full]
čm	<1>	*becsmérEL* [full]
fl	<2>	*caflAT, bifláZ* [full]
fr	<1>	*lófrÁL* [full]
sp	<2>	*diszponÁL, koszpitOL* [full]
st	<23>	*adjusztÁL, asszisztÁL, bizTAT?, dehonesztÁL, desztillÁL, egzisztÁL, engesztEL, gesztikulÁL, investÁL, készTET, magasztAL, marasztAL, molesztÁL, neheztEL?, osztOZIK, piszterkÁL, predesztinÁL, pusztUL, tapasztAL, tisztEL, vesztegEL, vesztEGET, vigasztAL* [full]
sk	<24>	*agnoszkÁL, alkalmazKODIK?, bámészKODIK, baszkurÁL, császkÁL, csimpaszKODIK, ereszKEDIK, fuoreszkÁL, foszforeszkÁL, hunyászKODIK, iszkOL, kapaszKODIK, mászkÁL, maszkírOZ, merészKEDIK, motoszkÁL, nyerészKEDIK, piszkÁL, poroszkÁL, prüszkÖL, reszkET, rugaszKODIK, tuszkOL, viszkET* [full]
sm	<3>	*eszmÉL, piszmOG, szöszmötÖL* [full]
sl	<1>	*észlEL* [full]
šp	<6>	*aspirÁL, hospitÁL, náspángOL, prosperÁL, respektÁL, tespED* [full]
št	<14>	*böstörKÖDIK, instÁL, istápOL, karistOL, kóstÁL, kóstOL, lósTAT, protestÁL, pustOL, restELL, rostokOL, **sisterEG**, testÁL, tüsténKEDIK* [full]
šk	<10>	*ágasKODIK, áskálÓDIK, bóbiskOL, diskurÁL, dúskÁL, kapiskÁL, miskárOL, paskOL, roskAD, viasKODIK* [full]

šm	<5>	***ismer**, ismétEL, kasmatOL, pusmOG, susmOG* [full]
šl	<4>	*búslaKODIK?, koslAT, pislANT, pislOG?* [full]
šh	<4>	*keshED, kushAD, peshED, poshAD* [full]
vl	<1>	*szívLEL?* [full]
zd	<3>	*?buzdÍT, mozdÍT, rezdÍT* [full]
zg	<4>	*birizgÁL, izgAT, izgUL, igazGAT* [full]
zv	<1>	*közvetÍT* [full]
zl	<2>	*?hízLAL, ?ízLEL* [full]
žd	<2>	*?pezsdÜL, pezsdÍT* [full]
žg	<2>	*babusGAT, rebesGET* [full]
žm	<1>	*prézsmitÁL* [full]
žl	<1>	?*vizslAT* [full]
mp	<16>	*amputÁL, csimpaszKODIK, extemporÁL, **hemperEG**, **hömpölyÖG**, imponÁL, importÁL, kámpicsorODIK, kalimpÁL, kompenzÁL, korrumpÁL, pumpOL, **sompolyOG**, szimpatizÁL, temperÁL, tempírOZ* [full]
mt	<1>	*nyomtat*? [full]
mb	<13>	*ambicionÁL, bömbÖL, **csámborOG,** döröbÖL, gémberEDIK, himbÁL, **imbolyOG,** kombinÁL, különbÖZIK, rombOL, tombOL, zörömbÖL, zsémbEL* [full]
mč	<2>	*csámcsOG, csemcsEG* [full]
mf	<1>	*somfordÁL* [full]
mž	<3>	*hemzsEG, karimzsÁL, körömzsÉL* [full]
ml	<5>	*emlEGET, emlÍT, kémlEL?, szemLÉL?, számLÁL?* [full]
nt	<31>	*argumentÁL, bántalmAZ?, bünTET, fontOL, frekventÁL, garantÁL, **henterEG**, intéZ, integrÁL, interjúvOL, internÁL, interpolÁL, interpretÁL, interveniÁL, intonÁL, kántÁL, kommentÁL, kontemplÁL, kurrentÁL, lamentÁL, orientÁL, parentÁL, plántÁL, plántÁL, poentírOZ, prezentÁL, reprezentÁL, tántorÍT, **tántorOG**, tünTET, vonTAT* [full]
ŋk	<16>	*bánKÓDIK, berzenKEDIK, bosszanKODIK, ellenKEZIK, ficánkOL, hetvenKEDIK, konkurrÁL, kunkOG, kunkorODIK, lankAD, pironKODIK, rimánKODIK, settenKEDIK, sopánKODIK, tehénKEDIK?, tüsténKEDIK* [full]
nd	<41>	*andalÍT, apprehendÁL, bandukOL, csendÜL, gondOL, gondOZ, gründOL, indÍT, indUL, kirándUL, komendÁL, kondicionÁL, kondoleÁL, kondUL, landOL, lendÜL, lendÍT, menDEGÉL, ondolÁL, penderEDIK, pendÜL, pocskondiáZ, rándÍT, rándUL, rekomendÁL, rendEL, rendÜL, sandÍT, skandÁL, spendírOZ,*

***sündörÖG**, szenderEDIK, szundÍT, szundikÁL, tendÁL, tündöklik, ugrándOZIK, undorÍT, vindikÁL, zendÜL, zsendÍT* [full]

ŋg	<21>	*ácsingÓZIK, barangOL, böngÉSZ, csatangOL, csilingEL, dangubÁL, döngÖL, engED, engedelmesKEDIK, engesztEL, flangÁL, kurjonGAT, langallik, marcangOL, náspángOL, pingÁL, pironGAT, rongÁL, szoronGAT, tángÁL, tengŐDIK* [full]
nt^{s}	<13>	*boncOL, cincÁL, cincOG, hancúrOZIK, hencEG, koncOL?, koncentrÁL, kuncOG, poncOL, ráncigÁL, szincerizÁL, toloncOL?, viháncOL* [full]
nč	<7>	*dancsOL?, incselKEDIK, **koncsorOG**, **kuncsorOG**, lincsEL?, pancsOL?, trancsírOZ* [full]
nf	<8>	*dezinficiÁL, inficiÁL, informÁL, konferÁL, konferÁL, konfirmÁL, konfundÁL, **ténferEG*** [full]
ns	<4>	*konszolidÁL, lanszírOZ, unszOL, vonszOL* [full]
nv	<8>	*invesztÁL, invitÁL, involvÁL, konveniÁL, konvergÁL, konvertÁL, konverzÁL, szenvED* [full]
nz	<4>	*inzultÁL, kompenzÁL, kondenzÁL, konzultÁL* [full]
nž	<2>	*avanzsÁL, revanzsÁL* [full]
nl	<2>	*kínlÓDIK?, színlEL?* [full]
nh	<1>	*inhalÁL* [full]
n^{y}t^{y}	<1>	*szontyolODIK* [full]
n^{y}d^{y}	<2>	*göngyÖL, kalangyÁL* [full]
n^{y}v	<1>	*senyvED* [full]
lp	<1>	*kolportÁL* [full]
lt	<10>	*álTAT, alterÁL, erölTET, inzultÁL, költÖZIK, konzultÁL, kultivÁL, mélTAT, öltÖZIK?, váltOZIK* [full]
lk	<17>	*álmélKODIK, **alkot**, **alkuszik?**, bővelKEDIK, bujálKODIK, emelKEDIK, foglalKOZIK, gyilkOL, hivalKODIK, kalkulÁL, kalkulÁL, lelkendEZ, sivalKODIK, találKOZIK, veselKEDIK, viselKEDIK, vulkanizÁL* [full]
lb	<1>	*kiakolbódÍT* [full]
ld	<8>	*áldOZ, foldOZ, fuldoklik?, haldoklik?, koldUL, nyeldekEL, **oldalOG**, üldÖZ* [full]
lg	<6>	*dédelGET?, dolgOZIK, hallGAT, latolGAT?, találGAT, vulgarizÁL* [full]
lt^{s}	<1>	*lefalcOL*? [full]
lv	<8>	*abszolvÁL, alvAD, devalvÁL, élvEZ, involvÁL, megnyilvánUL, olvAD, **olvas*** [full]

lm <10> *alkalmAZ, bántalmAZ?, cirkalmAZ, dalmahODIK, diadalmasKODIK, fogalmAZ, forgalmAZ, sugalmAZ, szolmizÁL, tartalmAZ* [full]

ln <1> ***volna*?** [full]

rp <10> *cirpEL, hörpINT, hörpÖL, horpAD, horpASZT, interpellÁL, interpolÁL, terpESZT?, torpAD, torpAN* [full]

rt <25> *abortÁL, amortizÁL, artikulÁL, csörTET, deportÁL, dezertÁL, disszertÁL, érteKEZIK, értelmEZ, értesÜL, exportÁL, fertőZ, firTAT, importÁL, irtÓZIK, kertEL, kolportÁL, konvertÁL, portÁL, szortírOZ, **tartóz**TAT, tartOZIK, **történik**, törTET, **zsörtöl**ŐDIK* [full]

rt^{y} <6> *fortyAN, fortyOG, hortyOG, szörtyÖG, szortyOG, vartyOG* [full]

rk <30> *acsarKODIK, babirkÁL, birKÓZIK, böstörKÖDIK, botorkÁL, butkOL, cirkÁL, cirkalmAZ, cirkulÁL, érKEZIK, férkŐZIK, fürkÉSZIK, hamarKODIK, horkAN, horkOL, iparKODIK?, ismerKEDIK?, kérKEDIK?, kurkÁSZ, markírOZ, pirkAD, piszterkÁL, pörkÖL, sarkALL, serkEN, serkENT, szemerkÉL, szerkESZT, szurkOL, zárKÓZIK* [full]

rb <4> *dorbézOL, turbékOL, verbuvÁL, zurbOL* [full]

rd <28> *cserdÍT, cserdÜL, csikordUL, csordÍT, csorDOGÁL, csordUL, csurdÍT, dördÜL, érdekEL, fordÍT, fordUL, gardírOZ, gördÍT, gördÜL, hördÜL, irdAL, kérdEZ, koordinÁL, mordUL, ordibÁL, perdÍT, perdÜL, serdÜL, somfordÁL, törd EL?, verdes?, zördÍT, zördÜL* [full]

rd^{y} <1> *meggárgyUL* [full]

rg <23> *argumentÁL, dezorganizÁL, dörgÖL, gargalizÁL, gurguláZ, háborGAT, hánytorGAT, hergEL, horgAD, horgOL, kapirgÁL, kergET, kergÜL, konvergÁL, kuporGAT, kurGAT, nyargAL, organizÁL, organizÁL, sürGET, tekergŐDZIK, vergŐDIK, zarGAT* [full]

rt^{s} <10> *exorcizÁL, farcinÁL, gürcÖL, hurcOL, percENT, percipiÁL, porcOG, sercEG, sercEN, varcOG* [full]

rč <1> *szürcsÖL* [full]

rf <5> *cserfEL?, köntörfalAZ?, morfondírOZ, perfektuÁL, perforÁL* [full]

rs <3> *forszírOZ, persziflÁL, perszonifikÁL* [full]

rš <5> *harsAN, harsOG, hersEG, mérsékEL, **társalOG**, versENG* [full]

rv <7> *hervAD, hervASZT, nyervOG, **örvend?**, sorvAD, sorvASZT, szervEZ* [full]

rz <4> *borzONG, borzAD, borzOL, toporzékOL* [full]

rž	<6>	*dörzsÖL, horzsOL, morzsOL, perzsEL, porzsOL, torzsalKODIK* [full]
rm	<14>	*alarmírOZ, dermED, dermESZT, dörmÖG, förmED, harmonizÁL, informÁL, körmÖL?, konfirmÁL, mormOG, mormOL, permutÁL, szármAZIK, zsurmOL* [full]
rn	<2>	*garnírOZ, internÁL* [full]
rn^{y}	<7>	*ernyED, ernyESZT, görnyED, kornyikÁL, szörnyED, vernyákOL, vernyOG* [full]
rl	<5>	*birLAL, érLEL?, porlAD?, súrlÓDIK?, törlESZT* [full]
rj	<15>	*burjánZik, erjED, erjESZT, gerjED, kimarjUL, kurjANT, kurjonGAT, őrjÍT, őrjÖNG, sarjAD, sarjadzik, sorjáZ, terjED, terjENG, terjESZT* [full]
rh	<1>	*archaizÁL* [full]
jp	<1>	*selypÍT* [full]
jt	<6>	*sajtOL, rejtŐZIK, fejtEGET, kojtOL,* ***kujtorOG****, folyTAT* [full]
jk	<3>	*hajkurÁSZIK, sulykOL, vájkÁL* [full]
jd	<11>	*bolydUL, bujdokOL?, bujDOS?, bujDOSIK?, gajdOL, jajdUL, kajdÁSZ, sajdÍT, sajdUL, sejdÍT, zajdUL* [full]
jg	<4>	*bolyGAT, stájgerOL, totojGAT, zajGAT* [full]
js	<2>	*hajszOL, majszOL* [full]
jš	<1>	*hajsókÁL* [full]
jv	<1>	*jajveszékEL?* [full]
jm	<2>	*bajmOL, lejmOL?* [full]
jn	<2>	*ajnárOZ, sajnÁL* [full]
jl	<5>	*bajlÓDIK, fájLAL?, fejlESZT, fejlŐDIK, hajlONG* [full]
jh	<1>	*megújhODIK*? [full]

(b) vsi-cluster=geminate <n=281>

p:	<24>	*csappAN, cseppEN, csippENT, cuppAN, cuppASZT, cuppOG, frappírOZ, huppAN, klappOL, koppAN, koppANT, lappAD,* ***lappang****, leppEG, röppEN, süppED, szeppEN, szippANT, töppED, toppAN, toppASZT, toppESZT, zsuppOL, zuppAN* [full]
t:	<14>	*csattAN, csattANT, csattOG, csettEN, csettENT, kattAN, pattAN, pattOG, ragadTAT, rettEG, rettEN, rettENT, suttOG, tetTET* [full]
t^{y}:	<16>	*fittyED, karattyOL, kettyEN, kettyENT, kettyINT, kottyAN, lottyAN, pittyED, pottyAN, pottyANT, rittyENT, suttyAN, suttyANT, szottyAN, tottyAN, zöttyEN* [full]

kː <35> *akkomodÁL, akkumulÁL, bukkAN, csökkEN, csökkENT, kukkAN, kukkANT, kukkOL, meghökkEN, meghökkENT, mukkAN, nyekkEN, nyikkAN, okkupÁL, pukkAN, pukkANT, pukkASZT, rekkEN, rikkAN, rikkANT, rokkAN, rukkOL, sikkAD, sikkANT, sikkASZT, smakkOL, szikkAD, szökkEN, szikkASZT, tikkAD, tikkASZT, vakkAN, vakkANT, zökkEN, zökkENT* [full]

bː <19> *csobbAN, döbbEN, döbbENT, dobbAN, dobbANT, fellebbEZ, gubbASZT, lebbEN, lebbENT, libbEN, lobbAN, lobbANT, meghibbAN, rebbEN, robbAN, robbANT, rokkAN, zsibbAD, zsibbASZT* [full]

dː <2> *edDEGÉL, idDOGÁL* [full]

d^{y}ː <5> *biggyED, biggyESZT, buggyAN, buggyASZT, roggyAN* [full]

gː <14> *aggASZT, aggAT, aggÓDIK, csüggED, csüggESZT, faggAT, guggOL, higgAD, higgASZT, nyaggAT, nyeggET, raggAT, szaggAT, szuggerÁL* [full]

t^{s}ː <13> *biccEN, biccENT, bliccEL, döccEN, koccAN, koccANT, koccINT, moccAN, pöccEN, pöccENT, ruccAN, spriccEL, truccOL* [full]

čː <9> *becsiccsENT, feccsENT, fröccsEN, fröccsENT, loccsAN, loccsANT, reccsEN, reccsENT, traccsOL* [full]

ǰː <1> *menedzsEL* [full]

fː <17> *böffEN, böffENT, buffAN, buffogAT, diffamÁL, kaffANT, kaffOG, laffOG, leffEG, összeröffEN, pöffEN, pöffENT, pöffeszkEDIK, puffAD, puffAN, puffASZT, röffEN* [full]

sː <27> *adresszÁL, asszimilÁL, bosszankODIK, bosszANT, csosszAN, csusszAN, deklasszÁL, disszertÁL, dresszírOZ, grasszÁL, masszírOZ, nyisszAN, nyisszANT, passzÍT, passzOL, pisszEG, pisszEN, prüsszEN, prüsszENT, prüsszÖG, sisszENT, szisszEN, szusszAN, szusszANT, tüsszEN, tüsszENT, tüsszÖG* [full]

šː <2> *tussOL, tessékEL?* [full]

zː <8> *duzzAD, duzzASZT, duzzOG, izzAD, izzÍT?, rezzEN, rezzENT, zizzEN* [full]

mː <9> *brummOG, cammOG, hümmÖG, kommentÁL, kommunikÁL, nyámmOG, nyammOG, stimmEL, zümmÖG* [full]

n^{y}ː <8> *dünnyÖG, fonnyAD, fonnyASZT, gunnyASZT, konnyAD, sunnyOG, szunnyAD, vinnyOG* [full]

lː <37> *alliterÁL?, apellÁL, **ballag**, billEG, billEN, billENT, brillírOZ, cizellÁL, csellENG, csillAN, csillANT, csillapÍT, csillapODIK, csillOG, desztillÁL, düllED, düllESZT, füllED, füllENT, füllESZT, fullAD, fullASZT, gyullAD, illAN, illEG-billEG, illEGET, illESZT,*

kallÓDIK, kellET, pallérOZ, pillANT, tallÓZIK, ***vállal****, villAN, villOG, villONG, züllESZT* [full]

r: <16> *cserrEG, csörrEN, csörrENT, csurrAN, csurrANT, durrAN, durrANT, durrOG, herrEG, kurrOG, surrAN, surrOG, virrAD, virrASZT, zörrEN, zörrENT* [full]

j: <4> *süllyED, süllyESZT, ujjONG, vijjOG* [full]

x: <1> *kehhENT* [full]

References

Aronoff, Mark. 1976. Word Formation in Generative Grammar. Cambridge, Mass.: MIT Press.

Bybee, Joan. 2001. Phonology and Language Use. Cambridge: Cambridge University Press.

Bybee, Joan, and Paul Hopper (eds.). 2001. Frequency and the Emergence of Linguistic Structure. Amsterdam: John Benjamins.

Chomsky, Noam and Morris Halle. 1968. The Sound Pattern of English. New York: Harper and Row.

Fudge, Erik C. 1970. Phonological structure and 'expressiveness'. Journal of Linguistics 6: 161–315.

Hay, Jennifer, Janet Pierrehumbert and Mary E. Beckman. 2003. Speech perception, well-formedness and the statistics of the lexicon. In: John Local, Richard Ogden and Rosalind Temple (eds.). Phonetic Interpretation. Papers in Laboratory Phonology VI. Cambridge: Cambridge University Press. 58-74.

Hayes, Bruce and Zsuzsa Cziráky Londe. 2005 Stochastic Phonological Knowledge: The Case of Hungarian Vowel Harmony. Ms. UCLA [http://www.linguistics.ucla.edu/people/hayes/Papers/HayesLondeHungarian_Aug05.pdf]

Itô, Junko and Armin. Mester. 1995. Japanese Phonology. In: John Goldsmith (ed.) The Handbook of Phonological Theory. Oxford: Blackwell. 817-847.

Kiefer, Ferenc (ed.). 1994. Strukturális magyar nyelvtan II. Fonológia. Budapest: Akadémiai Kiadó.

Kornai, András. 1986 Szótári adatbázis az akadémiai nagyszámítógépen, Műhelymunkák a nyelvészet és társtudományai köréből 2. 30-40.

Kiefer, Ferenc. 1998. Alaktan. In: É. Kiss Katalin, Kiefer Ferenc and Siptár Péter 1998. Új magyar nyelvtan. Budapest: Osiris Kiadó. 197-293.

McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. Linguistic Inquiry 12: 373–418.

Pierrehumbert, Janet. 1994. Syllable structure and word structure: a study of triconsonantal clusters in English. In: Patricia Keating (ed.). Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form. Cambridge: Cambridge University Press.

Rebrus, Péter. 2000a. Morfofonológiai jelenségek. In: F. Kiefer (ed.). Strukturális magyar nyelvtan 3. Morfológia. Budapest: Akadémiai Kiadó. 763–947.

Rebrus, Péter and Trón Viktor. 2002. A fonotaktikai általánosításokról: Kísérlet a magyar mássalhangzó-kapcsolatok nem-reprezentációs leírására. In: Maleczki Márta (ed.) A mai magyar nyelv leírásának újabb módszerei V. Szeged: Szegedi Tudományegyetem. 17-63.

Rebrus Péter and Trón Viktor. 2005. Re-presenting the past: Contrast and uniformity in Hungarian past tense suffixation. In: Kenesei István - Christopher Pinón - Siptár Péter (eds.). Approaches to Hungarian 9. Papers from the Düsseldorf Conference. Budapest: Akadémiai Kiadó.

Siptár Péter. 2002. Chapters from the segmental phonology of Hungarian. Academic doctoral dissertation. Budapest: MTA

Siptár, Péter and Miklós Törkenczy. 2000. The Phonology of Hungarian. The Phonology of the Worlds Languages. Oxford and New York: Oxford University Press.

Spang-Hannsen, Henning. 1959. Probability and Structural Classification in Language Description. Copenhagen: Rosenkilde and Bagger.

Tompa, József. 1961. A mai magyar nyelv rendszere I. Budapest: Akadémiai Kiadó.

Törkenczy, Miklós. 1994. A szótag. In: Kiefer 1994. 273–392.

Törkenczy, Miklós. 2000. Phonotactic grammaticality and the lexicon. In: Bartos Huba (ed.) Papers on the Mental Lexicon. Budapest: Research Institute of Linguistics, Hungarian Academy of Sciences. 211-230.

Törkenczy, Miklós. 2001. Phonotactic grammaticality and the lexicon. Acta Linguistica Hungarica 48: 137-153.

Törkenczy. 2004. The Phonotactics of Hungarian. Academic doctoral dissertation. Budapest: MTA.

Törkenczy, Miklós and Péter Siptár. 1999a. Hungarian Syllable Structure: Arguments for/against Complex Constituents. In: H. van der Hulst and Nancy. Ritter (eds.). The Syllable: Views and Facts. Berlin/New York: Mouton de Gruyter). 249–284.

Törkenczy, Miklós and Péter Siptár. 1999b. Vannak-e elágazó szótagösszetevők a magyarban? In: Kassai Ilona (ed.). Szótagfogalom — szótagrealizációk. Budapest: Linguistics Institute of HAS. 101–29.

Törkenczy, Miklós and Péter Siptár. 2000. Magánhangzó ~ semmi váltakozások a magyarban. Nyelvtudományi Közlemények 97: 64–130.

Trón Viktor and Rebrus Péter. 2000. Morphophonology and the hierarchical lexicon. In: Bartos Huba (ed.). Papers on the mental lexicon. Budapest, MTA Nyelvtudományi Intézet, 2000. 231-274.
Trón Viktor and Rebrus Péter: Morphophonology and the hierarchical lexicon. Acta Linguistica Hungarica 48: 2001: 101-135.
Vago, Robert M.. 1980a. The Sound Pattern of Hungarian. Washington: Georgetown University Press.
Vago, Robert M. 1989. Empty Consonants in the Moraic Phonology of Hungarian. Acta Linguistica Hungarica 39: 293–316.
Velcsov Mártonné. 1988. Alaktan. In: Rácz Endre (ed.). A mai magyar nyelv. Budapest: Tankönyvkiadó. 85-205.

Miklós Törkenczy
Department of English Linguistics ELTE
Theoretical Linguistics Programme
Institute for Linguistics
Hungarian Academy of Sciences
tork@nytud.hu