

Schrödinger's morpheme boundaries

Fejes, László

Hungarian Research Centre for Linguistics

fejes@nytud.hu

Affixes Symposium, University of Turku
17–18 August 2023

Project

- ▶ NKFI 139271
The role of paradigm structure in Hungarian morphology and phonology with typological comparisons
(led by Péter Rebrus, 2021–2025?)

Parallel historical developments

- ▶ **Proto-Uralic:** all stems are vowel-final
- ▶ **most of the descendants:** strong apocope: the emergence of consonant-final stems (at least in unsuffixed forms)
- ▶ **both Estonian and Hungarian:** the emergence of nominal forms in which the initial part is identical with the unsuffixed form and the final part is identical with the typical ending of the given paradigm slot, and there is a vowel between them.

Similar morphophonological structures in Estonian and Hungarian

Estonian			Hungarian		
NOM;SG	NOM;PL	meaning	NOM;SG	NOM;PL	meaning
<i>kana</i>	<i>kana-d</i>	'hen'	<i>kapu</i>	<i>kapu-k</i>	'gate'
<i>vaar</i>	<i>vaar-i-d</i>	'old man'	<i>vár</i>	<i>vár-a-k</i>	'castle'
<i>kaal</i>	<i>kaal-u-d</i>	'weight'	<i>kár</i>	<i>kár-o-k</i>	'damage'
<i>kaar</i>	<i>kaar-e-d</i>	'curve'	<i>őz</i>	<i>őz-e-k</i>	'roe'
<i>jaam</i>	<i>jaam-a-d</i>	'station'	<i>gőz</i>	<i>gőz-ö-k</i>	'steam'
<i>raam</i>	<i>raam-i-d</i>	'frame'	<i>hír</i>	<i>hír-e-k</i>	'news, fame'
<i>paar</i>	<i>paar-i-d</i>	'pair, couple'	<i>sír</i>	<i>sír-o-k</i>	'grave'
<i>saar</i>	<i>saar-e-d</i>	'island'	<i>íj</i>	<i>íj-a-k</i>	'bow'

Estonian: any vowel that usually occurs in non-initial syllables.

Hungarian: the possible vowels are limited (but not determined) by vowel harmony, but never high.

Similar structure, different analysis

Traditionally, these forms are segmented differently by the grammars of Estonian and Hungarian.

Estonian

NOM;SG	NOM;PL	meaning
<i>kana</i>	<i>kana-d</i>	'hen'
<i>vaar</i>	<i>vaari-d</i>	'old man'
<i>kaal</i>	<i>kaalu-d</i>	'weight'
<i>kaar</i>	<i>kaare-d</i>	'curve'
<i>jaam</i>	<i>jaama-d</i>	'station'
<i>raam</i>	<i>raami-d</i>	'frame'
<i>paar</i>	<i>paari-d</i>	'pair, couple'
<i>saar</i>	<i>saare-d</i>	'island'

Hungarian

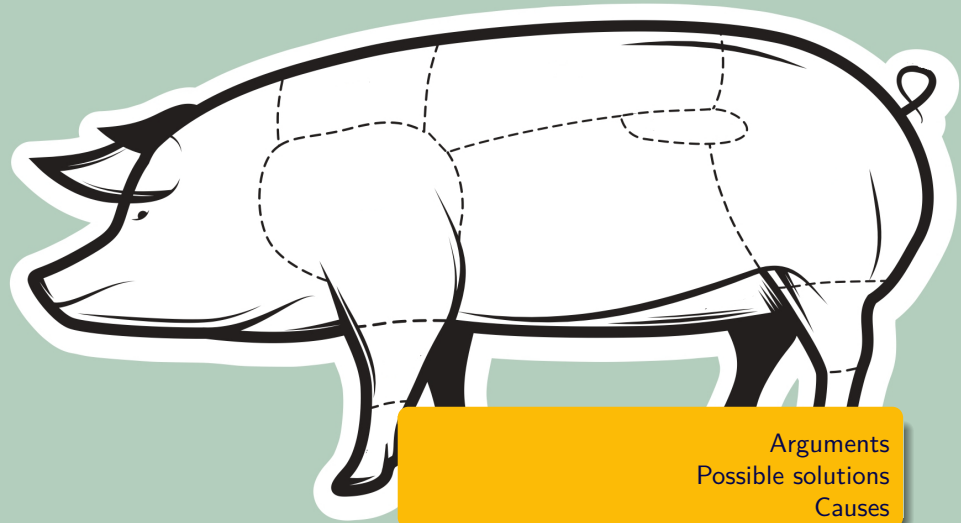
NOM;SG	NOM;PL	meaning
<i>kapu</i>	<i>kapu-k</i>	'gate'
<i>vár</i>	<i>vár-ak</i>	'castle'
<i>kár</i>	<i>kár-ok</i>	'damage'
<i>őz</i>	<i>őz-ek</i>	'roe'
<i>gőz</i>	<i>gőz-ök</i>	'steam'
<i>hír</i>	<i>hír-ek</i>	'news, fame'
<i>sír</i>	<i>sír-ok</i>	'grave'
<i>íj</i>	<i>íj-ak</i>	'bow'

Estonian type: stem-maximizing (SMax);

Hungarian type: suffix-maximizing (XMax)

Questions

- ▶ What are the arguments for and against the two kinds of segmentation?
- ▶ What would the correct analysis be?
- ▶ Do the two kinds of segmentation reflect some difference in the morphology of the two languages? Are there linguistic or extralinguistic reasons for the two traditions of segmentation?



Arguments
Possible solutions
Causes

Historical overview

- ▶ **Estonian:** never debated?
- ▶ **Hungarian:**
 - ▶ dominantly XMax;
 - ▶ **The Simonyi–Führer debate**
 - ▶ Simonyi (18??): SMax;
 - ▶ Führer (1881): against;
 - ▶ Simonyi (1881): the vowel does not belong either to the stem or the suffix: “linking vowel” (*kötőhangzó*, LV);
 - ▶ dominantly LV?
 - ▶ **The Antal–Berrár–Tompa debate**
 - ▶ Antal (1959) for XMax;
 - ▶ Berrár and Tompa (1960a) partially against (different analyses are possible);
 - ▶ Antal (1960) again for XMax;
 - ▶ Berrár and Tompa (1960b) for relativization
 - ▶ dominantly XMax (by time)
 - ▶ LV in education;
 - ▶ exceptions, e.g. Havas (1974) SMax.

Main arguments arguments

- ▶ **Against both:** too many allomorphs
 - ▶ SMax: too many stem allomorphs (Führer 1881: 310),
XMax: too many suffix allomorphs (Simonyi 1881: 373): so many forms of the vowel “that the language instinct could not consider these sounds belonging to the suffix.”
↔ Führer 1881 counts all kinds of vowels to the stem (*háza-m*, *házo-n*, *házu-nk*), ↔ either the stems or the suffixes must have many allomorphs
- ▶ **Against SMax:**
 - ▶ the suffix would determine the form of the stem (Führer 1881: 310)
↔ *utca* 'street' : *utca-i* 'street-ADJZ' : *utcá-ig* 'street-TERM';
 - ▶ inconsistency: *háza-m* 'house-POSS.1SG', but *ház-a* 'house-POSS.3SG' (Führer 1881: 311)
→ *koro-m* 'age-POSS.1SG', but *kor-a* 'age-POSS.3SG';
 - ▶ “odd” stems: *szomorúa-n* 'sad-MOD', *láb-úa-k* 'foot-ADJZ-PL', *buda-ia-k* 'Buda-ADJZ-PL' (Führer 1881: 311)
↔ oddity is relative;

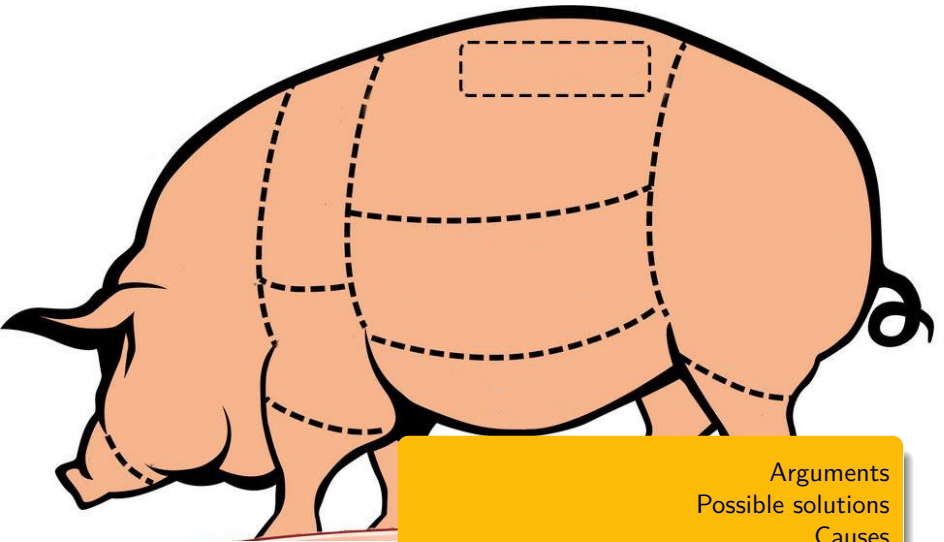
Main arguments arguments

► Against SMax:

- *ház-ak-at, kez-ek-et* vs. *háza-ka-t, keze-ke-t* 'house/hand-PL-ACC'
– *a/e* occurs after plural *k* only before accusative *t* (Antal 1959: 353)
← there is only one case suffix which shows the given type of alternations; before possessive suffixes, another plural marker is used;
- the primacy of the “dictionary” form: the nominal singular “basic” form is the basis of the comparison among the members of the paradigm (Tompá 1960a: 49, cf. Simonyi 1881: 372 on Estonian)
↔ *szavak* (: *szó*), *torkok* (: *torok*), *bírák* (: *bíró*), *falvak* (: *falu*), *nyarak* (: *nyár*), *csehek* (: [cse]) etc.? + Estonian grammars
- historical changes: *só* : *savak* 'salt/acid SG:PL' > *sav* : *savak* (**sava* : *savak*) 'acid SG:PL' (Tompá 1960b: 408)
**sava* : *savak* is impossible in Hungarian, the plural of **sava* would be *savák*
cf. hypothetical Estonian **soo* : *savad* > *sava* : *savad* (***sav* : *savad*);
- it is the suffix that requires the vowel (Laczkó 1997: 33)
↔ it is the stem that determines the quality of the vowel;

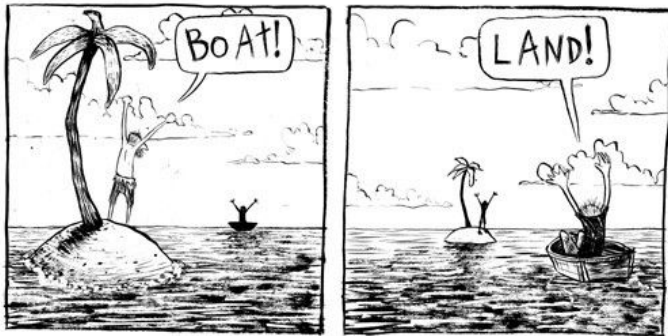
Main arguments arguments

- ▶ **Against the independent linking vowel:**
 - ▶ Every phoneme belongs to a morpheme and every morpheme has a meaning: as the linking vowel has no independent meaning, it cannot be an independent morpheme (Antal 1959: 352–353)
 - ↔ theory-dependent, cf. *álmorféma* (pseudo-morpheme; Faluvégi et al. 1994: 5, Szikszainé 1996: 160 etc.): segmentable, but without meaning – cf. intrusive (e.g. “hiatus-filling”) consonants/semivowels
- ▶ **Against XMax:**
 - ▶ The vowel quantity depends on the stem, additionally, on lexical, not phonetic factors and constant with a wide range of inflectional and derivational suffixes (Berrár 1960a: 46, Tompa 1960a: 49, cf. also Nádasy–Siptár 1989: 24–25)
 - ↔ the presence or the absence of the vowel depends on the suffix;
 - ▶ *ház-ak-at, kez-ek-et* vs. *háza-ka-t, keze-ke-t* ‘house/hand-PL-ACC’ the suffix alternations can be determined exclusively phonemically (not morphemically) in the latter case (Berrár 1960b: 407).



Arguments
Possible solutions
Causes

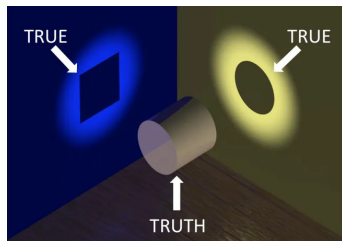
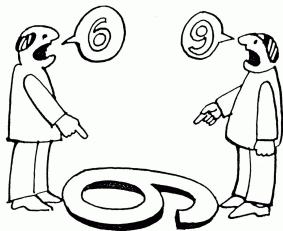
Trap 1: the distributional analysis



The vowel seems to belong to the suffix when paradigmatic slots with and without the vowel are compared:

- ▶ **Hungarian:** 'stalk, shank' *szár, szárnak, szárhoz, szárba* 'stalk' etc.
vs. *szárak, száram, szárad, száras* etc.
- ▶ **Estonian:** 'island' *saar, saart, saarde, saarte* etc.
vs. *saares, saarel, saareks* etc.

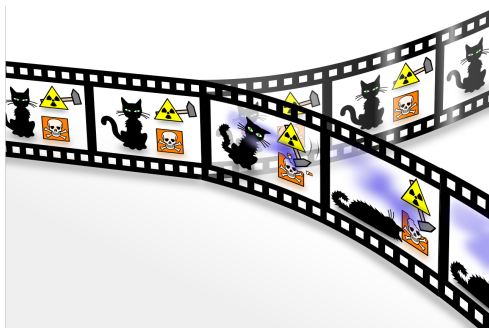
Trap 1: the distributional analysis



The vowel seems to belong to the stem when paradigmatic slots with the vowel are compared to each other or with similar paradigm slots of words ending in a vowel in their basic forms:

- ▶ **Hungarian:** ‘stalk, shank’ *szárok, száram, szárad, száras* etc.
(vs. ‘curd cheese’ *túrók, túróm, túród, túρός*)
- ▶ **Estonian:** ‘island’ *saares, saarel, saareks* etc.
(vs. ‘land’ *maas, maal, maaks* etc.)

Schrödinger's cat in the trap of distributional analysis



The morpheme boundary is in a kind of superposition, but it will be in a given place when an observation takes place (that is when the word form is contrasted to another one during the distributional analysis).

Trap 2: the concatenative approach

If a word is built up of distinct units (in the given case, a stem and suffixes), a segment is present in some forms of the paradigm but missing in others

- ▶ may belong to the preceding unit (stem);
- ▶ may belong to the following unit (suffix);
- ▶ may belong to neither the preceding or following element, that is may be independent.

If we do not stick to the concatenative approach, there are two additional possibilities:

- ▶ The segment may belong to both the preceding and the following unit (overlapping morphemes);
- ▶ The word should not be segmented at all.

Overlapping morphemes

Berrár 1960a: 44–45, 1960b: 406:

- ▶ There are morphemes, the boundaries of which are still/already [!!!] blurred;
- ▶ In Hungarian, the accusative case suffix is *-t*, the plural marker is *-k*: the vowel does not belong either to the stem or to the suffix;
- ▶ the distributional analysis does not determine morpheme boundaries in a univocal way: the existence of morphemes is undoubted, but the sharpness and consistency of their boundaries is not.

Overlapping morphemes

The fact that the quality of the vowel is determined by the stem but its appearance is determined by the suffix can be described in autosegmental phonology:

- ▶ stems end in floating (unanchored) vowel features;
- ▶ suffixes begin with a timing point unassociated with features;
- ▶ the vowel appears when the floating features can be associated to the timing point.

(Cf. Nádasy–Siptár 1989: 24–25)

Against overlapping morphemes

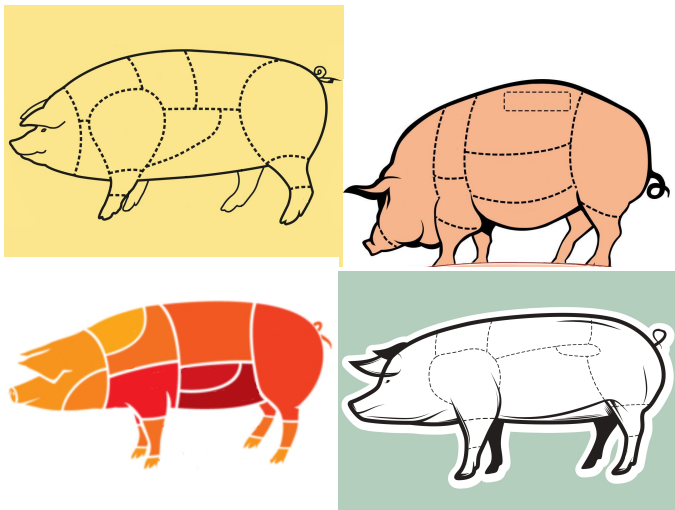
One morpheme (e.g. the stem) can impact a part of another morpheme (e.g. the suffix) not only at the morpheme boundary, but also inside the affected morpheme.

Hungarian vowel harmony

- ▶ **Phonological:** *ház-ban* 'house-INE' vs. *kéz-ben* 'hand-INE'
- ▶ **Morphological (lexical):** *sír-ban* 'grave-INE' vs. *hír-ben* 'news-INE'

The stem and the suffix morphemes are overlapping in the suffix vowel, but the consonant before it clearly belongs to the suffix. (Cf. discontinuous morphology in Semitic languages.)

Are there boundaries if there are no clear-cut boundaries?



The segmentation of pigs
according to websites of various Hungarian meat processing plants.

A passion for segmentation

Linguists tend to analyse words into morphemes even when there is no equivalence in form and function.

- ▶ (random¹) Slavic nominal “suffix” *-a*:
 - ▶ singular nominative for certain feminine (rarely masculine) lexemes;
 - ▶ singular genitive for certain masculine or neutral lexemes;
 - ▶ plural nominative for certain neutral lexemes;
 - ▶ plural accusative for certain neutral lexemes etc.
 - ▶ While the same function is fulfilled by other “suffixes” for other lexemes.
- ▶ (nominative) plural in German:
 - ▶ $-\emptyset$, $\ddot{\emptyset}$, *-e*, \ddot{e} , *-er*, \ddot{er} , *-n*, *-en*, *-s*
 - ▶ In what sense are these endings suffixes?

Such segmentations can be useful in practice (e.g. education), but they do not reflect the linguistic facts. They are not components of the language despite that they can be segmented in certain types of paradigms.

¹except for Macedonian and Bulgarian

Reasons for segmentation

- ▶ It is easy to formulate rules within concatenative morphology (“to generate the form X, add x to the stem”).
- ▶ It is relatively easy to find cases in any language (with morphology) where two forms differ only in their final (or initial) part and to say that the identical part is the stem and the differing parts are the suffixes (especially if there is a similar correspondence between the similar slots in several other paradigms).



Arguments
Possible solutions
Causes

A difference between Hungarian and Estonian

While the “linking vowel” never occurs independently (unfollowed by other phonemes, i.e. word-finally) in Hungarian, the vowel playing a similar role in Estonian does: it appears the singular genitive form.

NOM;SG	GEN;SG	NOM;PL	meaning
<i>kana</i>	<i>kana</i>	<i>kana-d</i>	'hen'
<i>kaal</i>	<i>kaal-u</i>	<i>kaal-u-d</i>	'weight'
<i>jaam</i>	<i>jaam-a</i>	<i>jaam-a-d</i>	'station'
<i>raam</i>	<i>raam-i</i>	<i>raam-i-d</i>	'frame'
<i>saar</i>	<i>saar-e</i>	<i>saar-e-d</i>	'island'

In an XMax analysis, not just genitive suffix allomorphes *-i*, *-u*, *-e*, *-a* (in addition to $-\emptyset$ etc.) should be supposed, but the nominative plural marker and almost all the case suffixes would begin with the same vowel. The problem also exists in Hungarian, but it is less salient due to the lack of the independent “linking vowel” and the limitedness of the vowel choice by vowel harmony.

Swings and roundabouts

The benefits of using the SG;GEN form as a stem in Estonian:

- ▶ usually, the same stem-final vowel is present in the non-concatenative so-called “short” ILL;SG and in many cases the of the PART;SG forms;
- ▶ based on the stem, the form of the so-called “*i* plural” can be well predicted: $a \rightarrow u/i$ (depending on the vowel in the initial syllable); $e \rightarrow i$; $u, i \rightarrow e$ (similarly to stems ending in vowels in NOM;SG);
- ▶ likewise, $u, a \rightarrow e$ before the comparative suffix in certain paradigm types (irrespective of the presence of the vowel in the NOM;SG form).

No similar phenomena are witnessed in Hungarian.

More drastic alternations

Additionally, Estonian has a lot of phenomena which could hardly be analysed as suffix alternations; therefore, the supposition of a stem allomorph is inevitable:

- ▶ *tuli* 'fire;NOM;SG' : *tuled* 'fire;NOM;PL';
- ▶ *teine* 'other;NOM;SG' : *teised* 'other;NOM;PL';
- ▶ *pime* 'dark;NOM;SG' : *pimedad* 'dark;NOM;PL';
- ▶ *habe* 'beard;NOM;SG' : *habemad* 'beard;NOM;PL';
- ▶ *liike* 'member;NOM;SG' : *liikmed* 'member;NOM;PL';
- ▶ *tuba* 'room;NOM;SG' : *toad* 'room;NOM;PL' etc.

Similar cases also occur in Hungarian, but in contrast to Estonian, all these types are unproductive and also less systematic:

- ▶ *nyár* 'summer;NOM;SG' : *nyarak* 'summer;NOM;PL';
- ▶ *ló* 'horse;NOM;SG' : *lovak* 'horse;NOM;PL';
- ▶ *mű* 'work;NOM;SG' : *művek* 'work;NOM;PL';
- ▶ *hű* 'faithful;NOM;SG' : *hűek* ~ *hívek* 'faithful;NOM;PL';
- ▶ *falu* 'village;NOM;SG' : *faluk* ~ *falvak* 'village;NOM;PL' etc.

Swings and roundabouts

A further benefit of rejecting concatenative morphology in GEN;SG is that based on the allomorph identical with the GEN;SG form, further forms (INE;SG, ELA;SG, ADE;SG, ABL;SG ALL;SG, ... NOM;PL) can be analysed concatenatively without additional complications.

The situation is similar in Hungarian, but there are less inflectional forms with the vowel in question, and the correspondence is less consistent:

- ▶ *vár* 'castle;NOM;SG' : *várak* 'castle;NOM;PL' : *várat* 'castle;ACC;SG'
- ▶ *kár* 'damage;NOM;SG' : *károk* 'damage;NOM;PL'
: *kárt* 'damage;ACC;SG'
- ▶ *hosszú* 'long;NOM;SG' : *hosszúak* ~ *hosszúk* 'long;NOM;PL'
: *hosszút* 'long;ACC;SG'
- ▶ *nyár* 'summer;NOM;SG' : *nyarak* 'summer;NOM;PL'
: *nyarat* 'summer;ACC;SG';
- ▶ *falu* 'village;NOM;SG' : *faluk* ~ *falvak* 'village;NOM;PL'
: *falut* 'village;ACC;SG' etc.

The Finnish connection

Finnish is closely related to Estonian but has kept the stem final vowels.
To describe Finnic languages in a similar way is rational.

In cognates, the vowel is present in all the forms:

- ▶ **Estonian** *vaar* 'old.man;NOM;SG' : *vaari* 'old.man;GEN;SG' :
vaarid 'old.man;NOM;PL'
- ▶ **Finnish** *vaari* 'grandpa;NOM;SG' : *vaarin* 'grandpa;GEN;SG'
: *vaarit* 'grandpa;NOM;PL'

The Finnish connection

In some cognates, there is vowel alternation: in the form corresponding to the Estonian vowelless form, another vowel is present:

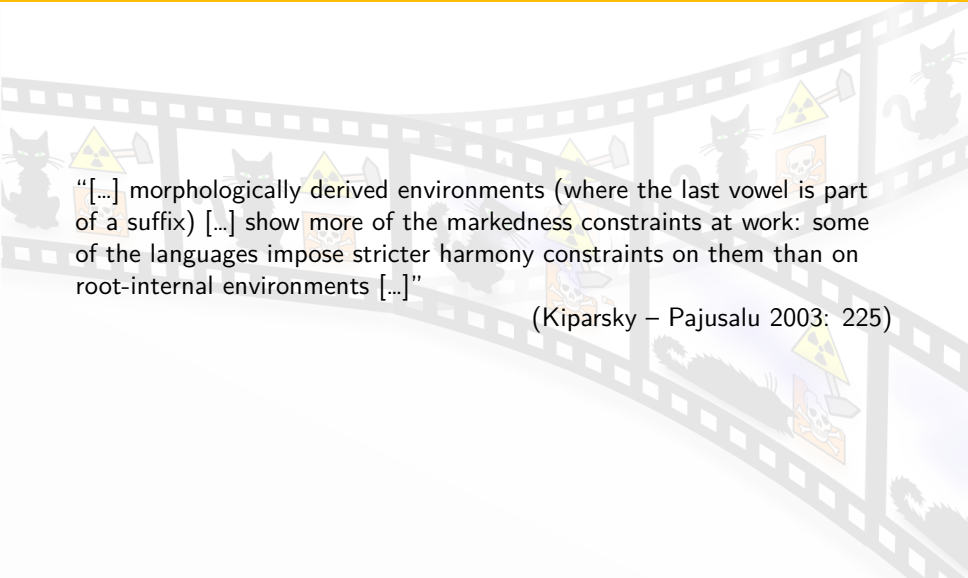
- ▶ **Estonian** *saar* 'island;NOM;SG' : *saare* 'island;GEN;SG'
: *saared* 'old.man;NOM;PL'
- ▶ **Finnish** *saari* 'island;NOM;SG' : *saaren* 'island;GEN;SG'
: *saaret* 'old.man;NOM;PL'

In Finnish grammars, stems with different vowels are analysed as different allomorphs of the same stem: it is supported by the fact that the types of such alternations are limited (*i* : *e*, *i* : *a/ä*).

Summary

- ▶ **What are the arguments for and against the two kinds of segmentation?** There are weaker and stronger arguments against both analyses, neither of them is ultimately convincing.
- ▶ **What would the correct analysis be?** There is no correct analysis, the best analysis would suppose overlapping segments or no segmentation at all.
- ▶ **Do the two kinds of segmentation reflect some difference in the morphology of the two languages? Are there linguistic or extralinguistic reasons for the two traditions of segmentation?**
In Estonian, stem allomorphs are more unpredictable based on the NOM.SG form than in Hungarian, but many oblique forms are much more predictable from the NOM.SG form, which differs only in the presence of an extra vowel in the stem type examined. The Estonian SMax analysis could be influenced by the grammars of Finnish, but it is not the main reason for the difference from the traditional Hungarian XMax analysis.

Why is segmentation so important?



“[...] morphologically derived environments (where the last vowel is part of a suffix) [...] show more of the markedness constraints at work: some of the languages impose stricter harmony constraints on them than on root-internal environments [...]”

(Kiparsky – Pajusalu 2003: 225)

References I

- ▶ Antal László. 1959. Gondolatok a magyar főnév birtokos ragozásáról. *Magyar Nyelv* LV. 351–357.
- ▶ Antal László. 1960. Mitől függ, hogy hány morféma? *Magyar Nyelv* LVI. 358–360.
- ▶ Berrár Jolán. 1960a. A magyar főnév birtokos ragozásáról. *Magyar Nyelv* LVI. 43–46.
- ▶ Berrár Jolán. 1960b. A morfémahatárok megállapításáról. *Magyar Nyelv* LVI. 465–467.
- ▶ H. Bottyánfy Éva – Horváth Mária – Korompay Klára - D. Máta Mária. 1976. *Bevezetés az egyetemi magyar nyelvészeti tanulmányokba*. Budapest: Tankönyvkiadó.
- ▶ Faluvégi Katalin – Keszler Borbála – Lackó Krisztina (szerk.). 1994. *Magyar leíró nyelvtani segédkönyv*. Budapest: Nemzeti Tankönyvkiadó.

References II

- ▶ Führer Ignác. 1881. Segéd vagy tőhangzó? *Magyar Nyelvőr* X. 310–312.
- ▶ Havas Ferenc. 1974. *A magyar, a finn és az észti nyelv tipológiai összehasonlítása*. Budapest: Akadémiai Kiadó.
- ▶ Kiparsky, Paul – Pajusalu, Karl. 2003. Towards a Typology of Disharmony. *The Linguistic Review* 20 (2-4): 217-241. DOI: 10.1515/tlir.2003.009
- ▶ Laczkó Krisztina. 1997. Tipikus grammatikai hibák általános és középiskolás nyelvtankönyvekben. 1. rész. *Magyar Nyelvőr* 121: 30–38.
- ▶ Lengyel Klára. 1995. Átmeneti kategóriák a nyelv különböző szintjein. *Magyar Nyelvőr* 119: 310–333.
- ▶ Simonyi Zsigmond. 1881. Segéd vagy tőhangzó? *Magyar Nyelvőr* X. 366–376.

References III

- ▶ Szikszainé Nagy Irma 1996. Meditáció az álmondatról. *Néprajz és Nyelvtudomány* XXXVII. 159—167.
- ▶ Tompa József. 1960a. A magyar lenév birtokos ragozásáról. *Magyar Nyelv* LVI. 48–51.
- ▶ Tompa József. 1960b. A morfémahatárok megállapításáról. *Magyar Nyelv* LVI. 467–468.

Photos

- ▶ Hentesnél evők baráti köre (Facebook)
- ▶ Index
- ▶ Kometa
- ▶ Böllér Kft.
- ▶ Pecsényés Húsbolt (Facebook)
- ▶ r/exmoron (Reddit)
- ▶ piccsy.com (Pinterest)
- ▶ Czirják Csilla (blog.hu)
- ▶ Christian Schirm (Wikimedia Commons)