

## Alien Words In Our Everyday Life (On The Need For An Estonian Database Of Structurally Alien Words)

**Merike Mägedi**

Institute of the Estonian Language

Department of Lexicology

Roosikrantsi 6

10119 Tallinn

Estonia

merike@eki.ee

### Abstract

The paper deals with some problems related to the formation of a database of structurally alien words. First, the criteria used by the Estonian linguists to distinguish what we call *võõrsõna* (< *võõras* 'strange, foreign, alien') from the rest of the vocabulary needs some revision, for genetic and temporal facts do not necessarily correlate with structural criteria.

As we know, once a sign system has become established, the adoption of new signs is not arbitrary. Evidently there is also certain regularity behind the necessity for a foreign-looking linguistic unit to be adopted into use now and then. The Database of Structurally Alien Words now prepared at the Institute of the Estonian Language (IEL) is meant to help the users of the Estonian language to find their way in the realm of such words. Timely and open as it is, the database is designed to provide exhaustive information on the functional behaviour of each item included.

### Introduction

Every progressive linguistic community seeks to make its means of expression as precise as possible. In this process, neologisms or derivatives are created, while some words are borrowed from other languages, as new notions need to be designated. The new words, in turn, need to be fixed in order to help the user to feel confident in their correct pronunciation, orthography and inflectional paradigm. The inclusion of a new word in a dictionary as if grants the word a license to live, just in the form it happens to have been fixed in.

Estonian linguists – at present – define the term *võõrsõna* as a **loanword**, which has been adapted to the language structure but partly, if at all, thus displaying some alien structural features such as the letters *f, š, z, ž*, the word-initial *b, d, g*, main stress in a non-initial syllable, unusual consonant clusters (*spr, ntg, pft, gt*, etc.), or *o* or a long vowel in a non-initial syllable. In my opinion, however, attention is due to all words of an alien structure, whatever their origin. This way, we can also handle some home-spun neologisms, which – if looking alien – will probably earn the label 'artificial word'. This was meant to emphasize that not every structurally alien word need be of a foreign origin. Such approach enables us to define the alien word ignoring the time-axis and to make a clear difference between loanwords and alien words.

As the reader will have noticed, I deliberately avoid the term 'foreign' (vs. 'genuine') word, as that term would be misleading for an English reader.

**The origin of Estonian word stems**

Hitherto no linguist has investigated the alien/non-alien ratio in the Estonian vocabulary. H. Rätsep (1983) has found out the ratio of loanwords vs. genuine stems in literary Estonian, excepting recent international loans. His results are as follows:

Genuine stems 46,67–60,93%

Stem loans 40,48–48,71%

Coined stems 0,94–1,01%

[Rätsep 1983] (Figure 1)

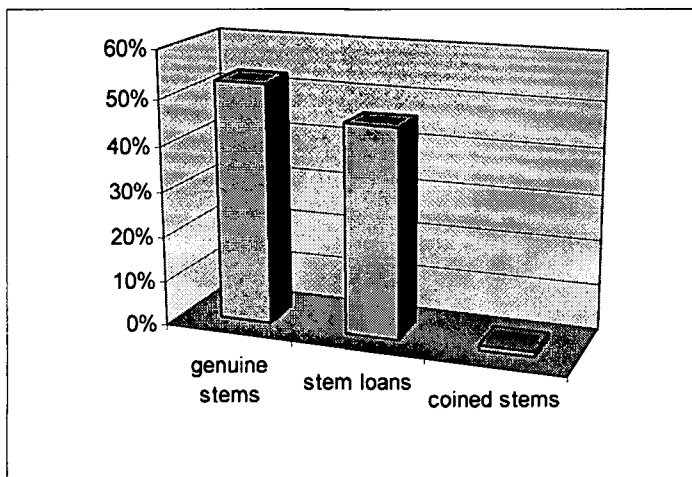


Figure 1: Borrowed vs. genuine Estonian stems

The above division is purely genetic. An alien word need not be automatically classified as a loanword. If it adapts oneself to the language, losing its alien features, it becomes a characteristically genuine word. A loanword can never be renamed a genuine word. This is why, while speaking of my database, I should rather avoid relying on the traditional English, German or any other terminology that differentiates between words by their origin and time of arrival.

**Definition**

In short, the problem of vocabulary classification can be approached in three ways:

- (1) Structural – alien vs. non-alien (genuine-looking) words
- Genetic – loanwords (newer and older) vs. genuine (incl. coined) words
- (2) Genetic – alien vs genuine words or non-alien words
- (Does it not also seem curious that we should speak of ‘borrowing’ and ‘loanwords’ referring to something that is never returned to the source?)
- (3) Genetic – loanwords vs genuine words
- either alien or non-alien in structure

For the purposes of my database I prefer the first variant, of course.

### The alien/non-alien word ratio in literary Estonian

The results of my own study of various texts published in the year 2000 reveal that in the Estonian literary language (as reflected in newspaper texts) there are 91% of non-alien words, and 9% of alien words (Figure 2)

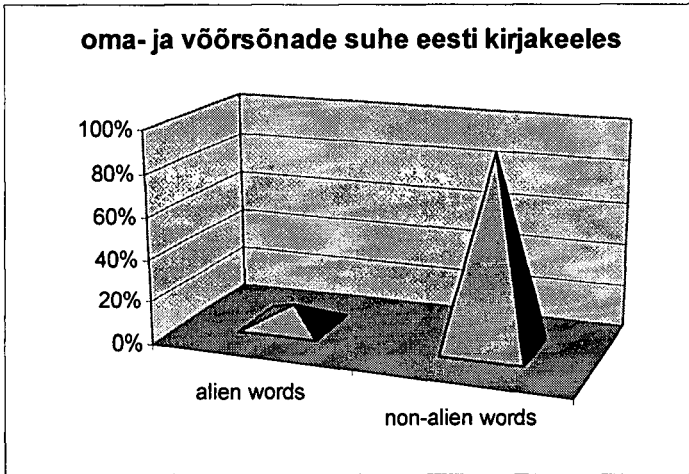


Figure 2: Alien vs. non-alien words ratio in the Estonian literary language

Although my study also embraced prose and scientific texts, newspaper texts are believed to be better in reflecting literary usage than any other type of written text.

In addition it occurred that the highest percentage (17.5%) of alien words are met in cultural press, whereas the lowest proportion (2%) of alien words is characteristic of poetry. (Figure 3)

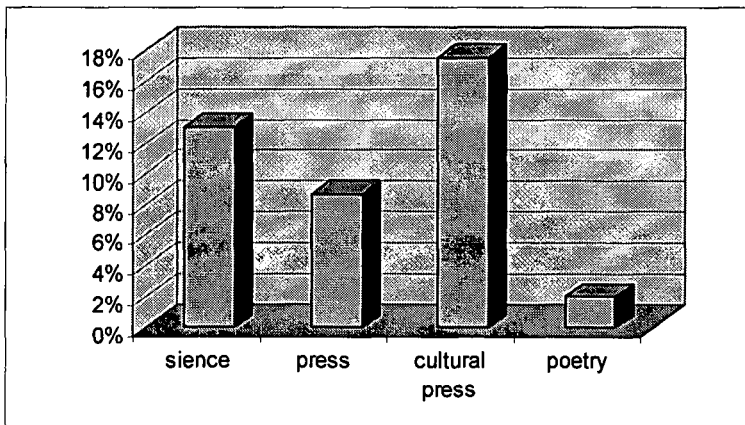


Figure 3: Percentage of alien words in various text types

### The Estonian literary norm

The normative advisory dictionary *Eesti Keele Sõnaraamat ÕS 1999* /Dictionary of the Estonian Language/ [ÕS 1999] generally advises the user to avoid words called *võõrsõna* (possibly characterised as alien words of foreign origin). Advice is given in four different ways. In the following examples A – alien word, non-A – non-alien word.

#### Structural recommendations

##### 1) A → non-A/A

- greider 'grader' *parem*: teehövel  
(alien features: word-initial *g* [non-alien compound is preferred])
- düün 'dune' *parem*: luide  
(alien features: word-initial *d*)
- retuusid 'leggings' *parem*: reiud  
(alien features: main word stress on a non-initial syllable)

The above examples demonstrate how the advisory dictionary handles those headwords, compounds, derivatives and phrases for which there are better ways of expression available. The preferable forms are referred to by the word *parem* 'better'.

##### 2) {A} → non-A, where {...} – word not to be used

This style of advising is stricter. The words not to be used are presented in braces and in a smaller print. There always follows an acceptable form of expression.

- {briifing} 'briefing' infotund  
(word-initial *b*, intra-word *f* [a hybrid word is preferred])
- {displei} 'display' kuvar  
(word-initial *d*, main word stress on a non-initial syllable)
- {arvutiseerima} 'computerise' arvutistama  
(alien derivative suffix [to be placed by a genuine] *-(i)seeri- ---> -sta-)*)

##### 3) non-A → A/non-A

Sometimes, though, an alien word is preferred to a genuine word as the former sounds semantically more precise and economical:

- asjaarmastuslik 'amateur' *parem*: amatöör
- nabapiirkond geogr. 'polar area' *parem*: polaarpiirkond

##### 4) A → A A

Sometimes the reference is the other way, supplementing a perfect expression by an extremely widespread, but not so good one introduced by *parem kui* 'better than':

- taideanatoomia 'plastic anatomy' *parem kui* plastiline anatoomia  
(main word stress and *o* in a non-initial syllable [preference belongs to a hybrid compound with at least one non-alien compound])

##### 5) A → A

- {massöör} 'masseur' masseerija
- testaator 'testator' *parem*: testeerija

(main word stress and long vowel in a non-initial syllable [for both words the original suffixes sound more foreign than the recommended combination of *-eeri-* and the genuine non-alien *-ja*])

### Semantic recommendations

Inappropriate meanings are referred to by *ei tähenda* 'does not mean'.

Sometimes the warning !HV.SOB ('seldom appropriate') is used:

konservator 'conservator; keeper' *ei tähenda*: konservatiiv 'conservative; Tory'  
 passioon 'passion' *ei tähenda*: granadill 'passion fruit'  
 simuleerima 'to simulate' *ei tähenda*: matkima 'to imitate'  
 elektroonika 'electronics' !HV.SOB 'elektronaparatuur' 'electronic equipment'  
 figuraalne 'figural' !HV.SOB 'kujundlik, piltlik' 'figurative'  
 figuur 'figure' !HV.SOB 'kujund' 'figure'

In order to regulate alien word usage we will have to use the above methods of reference in our database in progress. It is only natural that the Orthological Dictionary cannot include all words. In this respect a database can be more ambitious.

### IEL Database of Alien Words

An entry of our Database of Structurally Alien Words consists of a headword and attachment. Although the attached part differs for different headwords, it typically contains morphonological, grammatical, semantic, pragmatic and etymological information. Consequently, the database provides information not only on the orthography and meaning of the word, but also functional information on its language behaviour. The database is particularly valuable for its supply of abundant information on the use of this or that alien word and its non-alien equivalents, using glosses, labels, references and examples (include sentence citations).

The following examples are meant to provide a chance of comparison between the articles drawn from a couple of recent dictionaries [VL 2000; VS 1999] and our database of alien words.

**identifitseerima** (keskld *identificare*) e **identima** [-e'-], (midagi millegagi) samastama, samasust konstateerima v kindlaks tegema; *jur* kriminalistikas ja kohtumediisinis objekti tunnuseid võrdlema ja selle põhjal samasuse esinemist v puudumist järel dama; *sõj* sihtmärki tuvastama [VL 2000]

**identifitseerima** samastama; samasust kindlaks tegema; *jur* objekti tunnuste võrdluse põhjal samasuse olemasolu v puudumist järel dama; *loog* ühe eseme tunnuseid teisele, esimesega mõne muugi tunnuse poolest sarnanevale esemele omistama [VS 1999]

**identifitseerima** [identifits'eerima]

<v {mis/mille/mida} ({millega}) ({millena})> variant: identima

1 ära tundma, kindlaks tegema • *Ruth Benedict rõhutab, et kõikehõlmavaid kultuurimustreid on väga raske näha ja et neid peab õppima enda ümber igapäevases elus identifitseerima.* – M. Hinrikus

2 (väärtuste, normide, ideaalide põhjal) samastama, samaks pidama, (ennast) kuskile, kellegi hulka kuuluvana ära tundma, (oma) identiteeti määratlema *rühmitis ei identifitseeri ennast ühegi parteiga* [ÖS] • *Kõigele vaatamata pooldavad hispaanlased Euroopa Liitu ja identifitseerivad end eurooplastena.* – S. Krikk de Mateo; • *Eesti maastik ongi Eesti keha. Identifitseerin ennast*

*osana sellest kehast.. – T. Tammetalu --> vrd identiteet*

**3** JUR, INFO samasuse esinemise v puudumise põhjal ära tundma *laipa*  
*identifitseerima, andmelementi identifitseerima*

**4** LOOG ühe eseme tunnuseid omistama teisele, esimesega mõne muugi tunnuse  
poolest sarnanevale esemele

**5** SÕJ sihtmärki tuvastama [SÕJ, UVL]

▣ < KESKLD identificare < LD *idem* 'seesama' + *facere* 'tegema, sooritama' [VL,  
W91, KLUGE, LDEE]▣

/IEL Database of Alien Words/

## References

[Rätsep 1983] Rätsep, H., 1983. Eesti kirjakeele tüvevara päritolu, in: *Keel ja Kirjandus*, 10.

[VL 2000] Vääri, E. & Kleis, R. & Silvet, J., 2000. Võõrsõnade leksikon. (6. trükk). Valgus, Tallinn.

[VS 1999] Võõrsõnastik, 1999. TEA kirjastus, Tallinn.

[ÕS 1999] Eesti keele sõnaraamat ÕS 1999. Toimetanud Tiiu Erelt. Eesti Keele Sihtasutus, Eesti Keele Instituut, Tallinn.