The Criteria of a Loanword Etymology

Tartu 4.2.2016

Santeri Junttila

The anatomy of a loanword etymology

A loan etymology must consist of two parts, **lexical comparison** (rinnastus) and **claim of origin** (alkuperäväite).

Comparison: Fi. *haava* 'wound' ~ Lith. *šova* 'bolt, hollow'

Claim of origin: Fi. *haava* 'wound' ← Old Baltic

The criteria of a loanword etymology

I Formal criteria (muotokriteerit)
- A. Sound substitution (substituutio)
- B. Sound structure (äännerakenne)

II Semantic criteria (merkityskriteerit)
- A. Semantic corrispondence (merkitysvastaavuus)
- B. Semantic category (merkitysryhmä)

III Evidence of the source word (lähtökielikriteerit)

IV Distribution in the target language family (levikkikriteerit)
- A. Inclusive distributional criteria (positiiviset levikkikriteerit)
- B. Exclusive distributional criteria (negatiiviset levikkikriteerit)

V Evidence of the loanword layer (lainakerrostumakriteerit)
- A. Existence of the loanword layer (ontologiset kriteerit)
- B. Statistical probability (todennäköisyyskriteeri)

The criteria of a loanword etymology

I Formal criteria (muotokriteerit)

A. Sound substitution (substituutio)

B. Sound structure (äännerakenne)

II Semantic criteria (merkityskriteerit)

A. Semantic corrispondence (merkitysvastaavuus)

B. Semantic category (merkitysryhmä)

III Evidence of the source word (lähtökielikriteerit)

IV Distribution in the target language family (levikkikriteerit)

A. Inclusive distributional criteria (positiiviset levikkikriteerit)

B. Exclusive distributional criteria (negatiiviset levikkikriteerit)

V Evidence of the loanword layer (lainakerrostumakriteerit)

A. Existence of the loanword layer (ontologiset kriteerit)

B. Statistical probability (todennäköisyyskriteeri)

I A Sound substitution

All sound laws are **specific** for a certain *time* and *language* but **universal** inside this context.

This **Neogrammarian** discovery was the way out of *omnicomparativism*.

A sound law governs a **change** starting *phonetic* and turning into *phonemic* (phonological)

This **Structuralistic** discovery was the way out of *atomistic speculations*.

I A Sound substitution

All sound laws are **specific** for a certain *time* and *language* but **universal** inside this context.

Vilhelm **Thomsen** (1890) explained why Lith. ***ž**irnis* → Fi. ***h**erne* and Lith. *š**a**lna* → Fi. ***h**alla*, but Russ. *ж**а**ль* → Fi. *s**ä**äli* and Russ. *ло**ж**ка* → Fi. *lu**s**ikka*. The *conditions* had changed in between.

This **Neogrammarian** discovery was the way out of *omnicomparativism*.

Nikolai **Anderson** (1879) tried to compare Fi. *kirves* with Komi *čer* and Hu. *csőr*, since the change *k* > *č* is *globally common*.

A sound law governs a **change** starting *phonetic* and turning into *phonemic* (phonological)

This **Structuralistic** discovery was the way out of *atomistic speculations*.

I A Sound substitution

Sound laws are divided into **corrispondence rules** governing relationships between related languages and **substitution rules** governing representation of the sounds of the loan originals in the borrowed words.

Regular representation is a result of phonological substitution.

Fi. *halla* $<$ MPFi **šalna* $\leftarrow$ OBa **šalnā* $>$ Lith. *šalna*, where MPFi *š* $>$ LPFi *h* and MPFi *ln* $>$ LPFi *ll* are **regular changes governed by corrispondence rules** and OBa *ā* $\rightarrow$ MPFi *a* is a **regular substitution governed by a substitution rule.**

I A Sound substitution

Regular representation is a result of phonological substitution.

Substitution rules may tolerate variation due to changes in the conditions during a long-lasting language contact (phonetic changes in either of the languages included)..

Fi. *hall**a*** < MPFi **šalna* ← OBa **šalnā* > Lith. *šalna*, but

Fi. *hak**o*** < MPFi **šako* ← OBa **šakā* > Lith. *šaka*

..or due to differences in the phonetic inventory of the source and target language.

Fi. *t**a**rvas* < MPFi **tarwasi* ← OBa **tauras* > Lith. *tauras*

Fi. *t**o**rvi* < MPFi **torwi* ← OBa **taurē* > Lith. *taurė*

since OBa did not have an opposition of short *a* ~ *o*.

I A Sound substitution

Archaism check is needed to prove the loan etymology in cases where a sound change has lost a phonetic difference everywhere else but in an archaic language or dialect.

Fi. *lapio* 'spade' has been connected with PGmc **lapjō*, since Fi. *-io* could be either from PFi **-ijo* or **-ito*. But the Veps cognate *labid* can be only from **lapitV*, the Gmc etymology is wrong, and lapio must origin from OBa **lāpetā* > Lith. *lopeta*.

I A Sound substitution

Regular representation is the result of phonological substitution.

Acceptable exceptions to the regular representation, due to the moment of borrowing

- Morphological substitution
- Avoidance of homonymy (disputable)
- Morphological nativisation
- Semantical nativisation
- Iconical nativisation
- Etymological nativisation

Acceptable exceptions to the regular representation, due to later developments

- Wearing out of frequent words
- Expressive sound variation: phonaestemic, stem-internal and suffixal variation
- Etymological alienation (disputable)

Morphological substitution

Fi. *kiir*|*e* 'crown of head' ← Lith. *skyr*|*is* 'difference, class'

Fi. *synk*|*eä* 'gloomy' ← Lith. *sunk*|*us* 'heavy'

Avoidance of homonymy

Koivulehto käytti pyrkimystä homonymian välttämiseen argumenttina useitten lainaetymologioittensa puolesta: *suola* (ei ~ *suoli*, Koivulehto 1997a: 28), *vakka*, (ei ~ *vaka*, Koivulehto 2000a: 115), *ola* (ei ~ *vuoli*, Koivulehto 2000a: 118). Koivulehto (2009: 95–97) myös lakkasi pitämästä aiemmin (Koivulehto 1986b: 279–80) esittämäänsä *kestää*-verbin äännekehitystä < vksm **kenti*-(*tä-*) t. **keńć*-(*tä-*) ← ba **kentja-* (> lt *kęsti* : *kenčia* 'kärsiä, sietää, sallia') säännönmukaisena ehdottaessansa verbille *pettää* selitystä < vksm **pent-tä-* ← ba **spendja-* (> lt *spęsti*, prees. *spendžia* 'lõksu üles seadma'). Hän esitti selitykseksi, että odotuksenmukainen asu ***kettää* olisi johtanut häiritsevään homonymiaan toisen verbin *kettää* 'nylkeä nahka' kanssa.

Morphological nativisation

When a loanword represents another word class than its original, then the ending is a result of morphological nativisation.

Fi. *kirp|eä* 'tart, sharp' ← Lith. kirpti 'to become sour'

Semantical nativisation

Fi. *herhilä|inen* or *herhil|äinen* or *herhi|läinen* 'wasp' ← Lith. *širšuo*, *širšỹs*, *širšlys* 'hornet', cf. *ampiainen* 'hornet', *muurahainen* 'ant' etc.

Iconical nativisation

Fi., Kar. *harakka*, Est. *harakas* ~ Lydic, Veps *harag*, Vote *araga*, Liv. *aragys* 'magpie' ← Lith. *šarka* id.

Etymological nativisation

SEst. *ärk'* 'mite, tick' ? ← Lith. *erk*ė, Latv. *ērce* id.

Wearing out of frequent words

Expressive sound variation

synkeä ~ *synkkä* 'gloomy, grim'

halea ~ *haljas* ~ *haljakka* 'pale, watery (colour)'

Etymological alienation

Fi. *tiuris*, *tyyris* 'expensive' ← Sw. *dyr* id. < PSc **diuriz*

I A Sound substitution

Regular representation is the result of phonological substitution.

Acceptable exceptions to the regular representation, due to the moment of borrowing

- Morphological substitution
- Avoidance of homonymy (disputable)
- Morphological nativisation
- Semantical nativisation
- Iconical nativisation
- Etymological nativisation

Acceptable exceptions to the regular representation, due to later developments

- Wearing out of frequent words
- Expressive sound variation: phonaestemic, stem-internal and suffixal variation
- Etymological alienation (disputable)

I B Sound structure

- **Sound substance:** the etymology is the more trustworthy the more the word has sounds, especially infrequent sounds and sound combinations.
- **Phonological criteria** give a *terminus post quem*
- Fi. *porkkana* 'carrot' cannot be ← OBa, since its sound structure (geminate stop after a long syllable) is **too recent.**
- **Morphophonological criteria** give a *terminus ante quem*
- Fi. *varsi : varte-* 'handle' must be **at least as old as the end of the rules** *i* > *e* and *ti* > *si*, since both rules occur in it. Fi. *varsi* may be ← OBa **vardē*.

The criteria of a loanword etymology

I Formal criteria (muotokriteerit)
- A. Sound substitution (substituutio)
- B. Sound structure (äännerakenne)

II Semantic criteria (merkityskriteerit)
- A. Semantic corrispondence (merkitysvastaavuus)
- B. Semantic category (merkitysryhmä)

III Evidence of the source word (lähtökielikriteerit)

IV Distribution in the target language family (levikkikriteerit)
- A. Inclusive distributional criteria (positiiviset levikkikriteerit)
- B. Exclusive distributional criteria (negatiiviset levikkikriteerit)

V Evidence of the loanword layer (lainakerrostumakriteerit)
- A. Existence of the loanword layer (ontologiset kriteerit)
- B. Statistical probability (todennäköisyyskriteeri)

II A Semantic corrispondence

This is the most global but the least generalizable criterion, since in opaque cases we must operate with hypothetical semantic developments supported by parallels from any languages.

Fi. *peura* 'deer' is, according to Koivulehto, ← OBa **bendra-* > Lith. *bendras* 'common; comrade', since deer hunt used to be a common endeavour, and the catch was to be shared by the community. A parallel is Nsaa. *goddi* 'deer' = Fi *kunta* 'community, municipality', Mokša *końd'ä* 'comrade'.

A loan has often narrower meaning than the source word, since all words tend to polysemy but usually they get borrowed in one specific meaning only. But this is only a tendency with common exceptions.

II B Semantic category

This is a weak criterion alone, since some loans must be unique representatives of their sematic group in their loan layer, and since meanings may change. Still, when combined with other criteria, semantic category may give a *terminus post quem* in absolute chronology.

Fi. *pakana* 'heathen' cannot be ← OBa, since it belongs to the early Christian terminology, in which (morpho)phonetic properties, e.g. in *raamattu* 'Bible' and *risti* 'cross', prove an origin in a much more recent period of language history.

The criteria of a loanword etymology

I Formal criteria (muotokriteerit)
- A. Sound substitution (substituutio)
- B. Sound structure (äännerakenne)

II Semantic criteria (merkityskriteerit)
- A. Semantic corrispondence (merkitysvastaavuus)
- B. Semantic category (merkitysryhmä)

III Evidence of the source word (lähtökielikriteerit)

IV Distribution in the target language family (levikkikriteerit)
- A. Inclusive distributional criteria (positiiviset levikkikriteerit)
- B. Exclusive distributional criteria (negatiiviset levikkikriteerit)

V Evidence of the loanword layer (lainakerrostumakriteerit)
- A. Existence of the loanword layer (ontologiset kriteerit)
- B. Statistical probability (todennäköisyyskriteeri)

III Evidence of the source word

The existence of the claimed loan original in the source language should be proven. In praehistorical loans, the most reliable way is to show a large enough distribution in the related languages.

Fi. *kirstu* '(storage) chest' cannot be ← Lith. *karstas*, Latv. *šķirsts* id., since Lith. *karstas* ← Russ. ← Finnic, and Latv. *šķirsts* ← Est./Liv.

III Evidence of the source word

The existence of the claimed loan original in the source language should be proven. In praehistorical loans, the most reliable way is to show a large enough distribution in the related languages.

Could Fi. *salo* 'island; wilderness' be ← Lith. *sala* 'island', though this has no cognates in IE? Should we think about a common (substrate) origin instead?

III Evidence of the source word

Should we be able to prove that the claimed loan original in the source language was lexicalized?

Is Fi. *elki*, *elje* 'gesture; manner' ← Lith. *elgtis* 'to behave', though this has no nominal derivatives suitable as a direct loan original?

Sometimes these questions get an answer in the target language group: cf. Veps *elgeta* 'to understand'.

III Evidence of the source word

Should we be able to prove that the claimed loan original in the source language was lexicalized?

Are Fi. *puhdas* 'clean' ← Lith. *puošta-*, cf. *puošti* 'to decorate' and Fi. *kypsä* 'ripe, mature; done (of food)' ← Lith. *kepta-*, cf. *kepti* 'to roast, bake', though these are participle forms?

III Evidence of the source word

Should an indirect proof of an original be accepted?

Is Fi. *jutta* 'yoke strap' ← OBa, cf. Skr. *yuta-* id.??, though there are no direct cognates in Balto-Slavic languages?

IV Distribution in the target language family

- Inclusive distributional criterion is a relative *terminus ante quem*: if a word A of lg X has a real cognate in a related lg Y, then the word A must not belong to any layer later than the separation of X and Y. A *real cognate* means here that it cannot be proven a result of later borrowing between X and Y (by formal and semantic criteria).

IV Distribution in the target language family

- Exclusive distributional criteria are very weak, since, in principle, any word may disappear from any language. If a word B in lg X has no cognate in related lg Z, it proves just that the word B *may* belong to a layer later than the separation of X and Z.

IV Distribution in the target language family

Distribution is not a primary criterion but a result of previous etymological research. Thus, cognates in related languages are not a priori to be preferred to a loan origin, but formal and semantic criteria must be used to evaluate the concurrent etymologies.

Fi. *hara*, *harmaa*, *hihna*, *hiiva*, *kataja*, *käki*, *laiva*, *lapio*, *liiva*, *metsä*, *muli*, *palvoa*, *puuro*, *suka*, *teeri*, *temmata*, *varsi* and *varsta* have all got both a OBa loan etymology and proposed FP, FU or Ur cognates. In all cases, first an investigation of historical phonology and semantics has solved the case.

V Evidence of the loanword layer

- Existence of the loanword layer

How many loanwords (of each type) is needed to prove the existence of a stratum of borrowings?

- Statistical probability

Is Fi. *ranta* 'beach, shore' ← Lith. *krantas* id. or ← Sw. *strand* id.?