

## **The Treatment of Compound Words in a Language Planning Dictionary**

### **Abstract**

This paper deals with certain problems connected with the presentation of compound words in the Estonian Language Planning Dictionary. It is meant to explain how a certain group of words (such as compounds) has been harnessed to serve the purpose of a new kind of dictionary of modern literary Estonian, which is language planning.

This paper will provide a short glimpse into the two-volume Language Planning Dictionary (LPD) now in progress at the Institute of the Estonian Language, from the authors' perspective. The project was begun in 1988. In 1992, when Euralex first heard of the dictionary from the initiator of the project Henn Saari (Saari 1992) and Tiina Leemets (Leemets 1992), the team of four (with Tiiu Erelt as editor-in-chief), had managed to complete more than half of the manuscript. As of now the compilation has covered all but the final few letters of the alphabet.

The LPD is generally regarded as the successor to *Õigekeelsus-sõnaraamat*, the Orthological Dictionary (OD) the most recent edition of which dates back to 1976 (edited by Rein Kull and Erich Raiet). According to the Compiler's Guide, the LPD is targeted to a wide range of users, whoever needs to write, edit, correct or simply read texts in Estonian.

### **Principles underlying the selection of compounds in headword status**

The headwords of the LPD are based mostly on those of the OD, but their number has been considerably reduced. Fewer foreign words and special terms are included in the LPD, so as not to duplicate either the Lexicon of Foreign Words or the numerous terminological dictionaries recently published or in progress. As for compound words, their number in the OD is quite impressive. In the LPD however, their number is kept at a level which enables the user to learn how to form his or her own compounds. According to preliminary estimates, the LPD should have slightly less than half of the 115,000 headwords contained in the 1976

edition of the OD. However, since LPD entries are longer, the original one-volume design has expanded to at least two volumes.

A large number of compound words are presented in the OD. Most word combinations included are those often mistaken for one word. Since 1976 however, the rules for compounding have been changed, leaving more up to the writer to decide. Those changes are also reflected in the LPD. For example, LPD headwords include word-combinations, while their usage is explained by means of examples and comments.

*Example 1.*

**selja/taga, selja taga.** Pikk öö on seljataga v selja taga (*möödas*). Seljataga v selja taga kutsuti tüdrukut Karva-Mariks (*tagaselja*). *Vrd otseses tähenduses: Keegi köhatas selja taga*

Headwords are in bold type, examples in Roman type, while italics is reserved for equivalents and comments. A slash serves to separate compound components. The last example *Keegi köhatas selja taga* 'Behind his back someone cleared his throat' together with the preceding comment serve to explain that when used in its literal sense, the word combination should be written separately. However the sentences *Pikk öö on seljataga v selja taga* 'The long night is over'; *Seljataga v selja taga kutsuti tüdrukut Karva-Mariks* 'Behind her back the girl was called Hairy-Mary' exemplify various figurative senses of the headword, in which case compounding is allowed, albeit not required.

The LPD includes some phrasal verbs and expressive verbs, which can be written as one word in certain circumstances, whereas their derivatives, which are normally written as one word, are included in the entry of the corresponding verb (if not in headword status). E.g. the entry *sisse rändama* 'immigrate' includes such derivatives as *sisseränd* and *sisseränne* 'immigration', while the entry *sisse vedama* 'import' also manifests *sissevedaja* 'importer' and *sissevedamine* 'importation'. However the derivative *sissevedu* 'import', which appears as the first component of many compounds, has been given headword status.

From the point of view of the LPD, OD compounds fall into three categories: 1) compounds which will have their own entry, 2) compounds to be presented as examples within another entry, and 3) compounds to be left out of the LPD. Replacing the latter are of course many compounds adopted into use after 1976. Headword status is given to compounds which are more complex. This necessary additional information may appear as: 1) morphological data (if, in exceptional cases, all components of the compound are inflected), 2) labels of style, special field etc., 3) synonyms or explanations of definition, 4) examples of

usage to illustrate the meaning or some morphological aspect of the word, 5) compounds with the headword appearing either as the first or the final component, with the selection being linguistically, not by principles of definition, 6) derivatives of the headword, 7) cross-references. Each of the criteria mentioned justifies giving the compound in question a separate entry, while in many cases several conditions are met.

If a headword has an inadvisable equivalent of one or several words, a reverse entry is usually also provided, in which the headword is not recommended, but the equivalent is. If we did not present the inadvisable (“negative”) expression in the headword position, many users would find it difficult to locate the “positive” equivalent, as the inadvisable variants are often quite widespread and thus may be more well-known than the recommended variants. Therefore we have not been sparing with negative examples, but also take care not to leave them without positive counter-examples.

### Compounds to be presented as examples within another entry

A large group of LPD compounds are presented not as headwords, but as examples. These are compounds for which sufficient material has not yet been collected to justify a separate entry, yet are too significant to be left out of the LPD. Such compounds are in turn split into examples of the base component entries and attributive component entries. According to our own Compiler’s Guide, the selection of compound word examples should illustrate 1) the different types of compounding (typical forms of the attributive component) such as genitive, nominative, consonantal and truncated stems; and 2) groups by definition.

#### *Example 2.*

**saadetis** <12: -e>. Kauba/, posti/, raha/saadetis. Kiir/, liht/, luna/, ristpael/, täht/, väärt/, väike/saadetis

#### *Example 3.*

**saar 2.** <21: saare, saart> *veega piiratud maismaa*. Jõe/, järve/, mere/, ookeani/, soo/, kalju/, korall/, pool/, rõngas/saar *atoll*. Kõrbe/saar *oaas*. Ohutus/saar (*tänaval*). Keele/saar

The angular brackets following the headword contain morphological information such as the inflectional type number and inflected forms. For example 2 the compounds are grouped by type of compounding, so that

in the first group the headword *saadetis* 'something being sent' is the base for genitive attributes (*kaubasaadetis* 'consignment', *postisaadetis* 'mail', *rahasaadetis* 'remittance'), while the other group contains examples in which the same base word takes attributive stems either in the nominative or truncated form (*kiirsaadetis* 'quick dispatch', *lihtsaadetis* 'surface mail', *lunasaadetis* 'C.O.D. parcel', *ristpaelsaadetis* 'something sent by book post', *tähtsaadetis* 'registered letter (or packet, etc.)', *väärtsaadetis* 'insured post', *väikesaadetis* 'small dispatch'. In example 3 however, the compounds are grouped by meaning: *jõesaar* 'holm', *järvesaar* 'lake island', *meresaar* 'sea island', *ookeanisaar* 'oceanic island', *soosaar* 'bog island', *kaljusaar* 'rocky island', *korallisaar* 'coral island', *poolsaar* 'peninsula', *rõngassaar* 'atoll'; *kõrbesaar* 'oasis'; *ohutussaar* 'refuge, safety island'; *keelesaar* 'linguistic enclave'.

The preceding exemplary principle has a direct bearing on the authors' ambition to reduce the number of LPD compounds, since it would not be logical or possible to fit all Estonian compounds into a single dictionary. Instead, we are trying to provide examples with which the user of the LPD can construct new compounds. E.g. the Orthological Dictionary has such headwords as *sitsikleit* 'cotton dress', *sitskardin* 'cotton curtain', *sitskleit* 'cotton dress', *sitspluus* 'cotton blouse', *sitspõll* 'cotton apron', *sitsriie* 'cotton print', *sitsrätik* 'cotton kerchief', *sitsseelik* 'cotton skirt'. The LPD entry of *sits* 'cotton print' contains the compounds *sitskleit*, *sitsriie* and *sitsrätik* along with a note, that in colloquial language the genitive attribute *sitsi-* is quite popular yielding such forms as *sitsikardin*, *sitsikleit*, *sitsiriie* and *sitsirätik*. The compounds *sitspluus*, *sitspõll* and *sitsseelik* have been omitted from headwords as well as from examples as they are formed analogously to *sitskleit*, comment sustained (i.e. usually the attributive stem joins in the nominative case form, thus yielding *sitspluus*, *sitspõll*, *sitsseelik*, but in colloquial use genitive compounding is more common: *sitsipluus*, *sitsipõll*, *sitsiseelik*). The words *sitsseelik* 'cotton skirt' and *nahkseelik* 'leather skirt' can also be found among the examples of *seelik* 'skirt', but not the colloquial form *sitsiseelik*.

Consequently, if a compound does not occur among LPD headwords, it may be found in the entries of either the base component or the attributive component, but it need not occur in the LPD (or any dictionary) at all, despite being a common Estonian compound. Following the principles of compound treatment set out in the compilation guidelines, the authors of the dictionary must constantly consider the user's perspective, trying to imagine what could evoke the need to consult the dictionary and where he or she is most likely to look for the word, so that the right words will be in the right places.

On the one hand, if as many words as possible are included, there is the danger that the compounds will simply be distributed among entries, in an attempt to ensure that they are registered somewhere. It is perhaps the abundance of compounds presented in the OD that has misled users to believe that if a compound cannot be found in a dictionary, it simply does not exist, i.e. it should be treated as a word combination. On the other hand, with an excess of examples and additional compounds, some entries are bound to grow unduly long, which once again makes the dictionary awkward to use. This has already been noticed when using our own LPD manuscript.

At present all words considered essential for language planning purposes are presented in the LPD as headwords in strict alphabetical order. The remaining compounds occurring only among the examples are additional, and should be taken as such, without expecting them to include every compound not found among the headwords. Simultaneously, headword status does not exclude the recurrence of the same compound among additional examples, and vice versa. Although such repetition enlarges the volume of the dictionary, we are prepared to accept the consequences if it cases the word-accessibility. Therefore we often prefer repeating to cross-referencing, as in the following example:

*Example 4.*

**maja/pidamis/seep.** Tualett- ja majapidamisseeep, *parem* silma- ja pesuseep

**pesu/seep,** *parem kui* maja/pidamis/seep. Pesu- ja silmaseep

**silma/seep,** *parem kui* tua.let't/seep. Silma- ja pesuseep

**tualet't** <20: -leti, -letti> (*naiste*)riietus; riietumine; tualettlaud; klosett ja pesuruum. Kallites tualettides daamid. Teeb hommikust tualetti. Balli/, õhtu/tualett. Tuaelett/ruum = tualett, /laud, /paber, /seep, *parem* silmaseep, /vesi. Tuaelett/ = tualeti/tarbed

**seep** <20: seebi, .seepi>. Pesu/, silma/seep, *parem kui* majapidamis/, tualett/seep

In this way we hope that the reader would by no means fail to see that *silmaseep* is a far better word for scented soap than *tualettseep*, while *pesuseep* is recommended for referring to laundry soap instead of *majapidamisseeep*. As is obvious from the above-mentioned examples (except the last one, in which the compounds *pesuseep* and *silmaseep* appear as examples of their base word entries), the repetition of a headword among examples of the same entry, together with an antonym (or related word in other respects), as a counter-example, is a common

practice in the LPD. This helps avoid cross-referencing, especially to entries which lie at the other end of the dictionary.

### References

- Leemets, T. 1992. "A link between language planners and language speakers: a dictionary"; in: *Proceedings Euralex '92*, pp. 479–484.
- Saari, H. 1992. "Einsprachiges Handwörterbuch leserfreundlich"; in: *Proceedings Euralex '92*, pp. 503–510.