

Compounding in Esperanto*

Wim Jansen

ACLC, University of Amsterdam

This study is the third in a series on words and word building in Esperanto. Its predecessors addressed the parts of speech of the language and its derivational system. Here, compounding is discussed, whereby the analysis is restricted to compounds with two lexical components. Like derivation, compounding is described as a process which is hosted in the lexicon. Most compounds are found to be determinative and endocentric. The language also features exocentric copulative and possessive compounds. A critical analysis of epenthetic or bridging vowels inside compounds is given. The approach that is followed is inspired by techniques from Functional Discourse Grammar and is entirely novel compared with traditional linguistic descriptions of Esperanto.

1 Introduction

The first project of Esperanto was published in 1887 (Zamenhof 2004 [1887]), and the basic grammar and vocabulary of the language were baselined in 1905 (Zamenhof 1963 [1905]).¹ Since then, Esperanto has been used uninterruptedly in a growing variety of applications. Developments in the language out of the 1905 baseline have never been imposed by its author or any other authority, and most of them continue to occur spontaneously. They manifest themselves in the spoken and written day-to-day language, and are registered in the literature, prescriptive grammars and dictionaries which are published by the Esperanto community.² Kalocsay & Waringhien (1985), Waringhien (2005), and Wennergren (2005) are among the most authoritative works currently in use.

Major lexical developments have taken place in Esperanto since its early days. This paper describes one of them: word building by compounding. Compounding is addressed in just one single rule in the Russian edition Zamenhof (2004 [1887]) and in its subsequent four translations in Polish, French, German and English. This rule is reproduced as Rule 11 in the grammar

* The author thanks Federico Gobbo for his review of a draft of this paper, and for the comments provided.

¹ Lejzer Markovič Zamenhof (Białystok 1859 – Warsaw 1917), author of Esperanto.

² Rather than as a ‘speech community’, the Esperanto-speaking community should be referred to as a ‘community of practice’. In this short study, I will use the simple ‘community’.

section of Zamenhof (1963 [1905]), and is illustrated by a few cases in § 27 of the exercise section of the same book and, indirectly, by compounds used by the author in its foreword. Besides derivation, compounding is a powerful tool in Esperanto to coin new words. Derivation was addressed in detail in Jansen (2016a). This complementary study is concerned with compounds. The way compounding in Esperanto is explained as a lexical process follows the same approach that was taken in the earlier work on parts of speech and derivation, and is inspired by Functional Discourse Grammar (FDG). A brief recapitulation of this is given below, as it may serve as a concise introduction to the subsequent sections.

In the Esperantological literature, the term *radiko* ‘root’, is traditionally used to denote a morpheme with a lexical content that can occur as the sole lexical component of a word. I will maintain the translation ‘root’ in this study. Those elements which are commonly called *afikso* ‘affix’, *prefikso* ‘prefix’ and *sufikso* ‘suffix’ are limited in this study to dependent lexical morphemes that are used in conjunction with a root.³ Non-lexical suffixes I will refer to as ‘endings’. This way, I hope to achieve maximum coherence with the Esperantological traditions and with the general linguistic literature at the same time, realizing that this is not in compliance with the specific FDG terminology presented in Hengeveld & Mackenzie (2008: 404).

In Jansen (2013), it is argued that basic content words in Esperanto can be described as built on roots in an originally flexible system of parts of speech. These roots are associated with lexemes or heads of lexical properties, designating one of the following semantic categories: Individual (x), State-of-Affairs (e), Propositional content (p), Location (l), Time (t), Manner (m), Quantity (q), Reason (r), Instrument (ins) and Dependents (f). Given the one-to-one translation that Zamenhof gives of each single root in Esperanto as a syntactic word in each of the five reference languages, I propose to adopt in this word-oriented study more user-friendly names for (p) and (f), i.e., Abstractum for (p), primarily for nominal encoding, and Qualificational (qual) and Situational (sit) for the Dependents (f), following their primary word realizations as adjectives and verbs respectively.⁴

An Esperanto root is not ready for operation in the syntax. It takes the ending *o* to become a noun N, *a* as an adjective A, *i* as a verb V in the infinitive mood and *e* as an adverb Adv. Consequently, the parts-of-speech system is fully

³ This is how *prefiksoj* and *sufiksoj* were presented in the earliest Esperanto sources. Admittedly, many of them underwent a process of lexicalization over the years, and gave rise to a series of pseudo-roots, but this does not contradict their dependent origin.

⁴ ‘Qualificational’ should be self-explanatory. ‘Situational’ is related to an older denomination in FDG of configurational properties, which contain the predication frames. In Esperanto, verbs are the compulsory nuclei of predicates.

specialized at the syntactic level, with obligatory morphological marking of each root as it is used as a word in a slot reserved for one of the four open classes, e.g., the nouns *pluv-o* ‘rainfall’ built on PLUV⁵ (e) and *esper-o* ‘hope’ < ESPER (sit); adjectives like *bel-a* ‘beautiful’ < BEL (qual) and *amik-a* ‘friendly’ < AMIK (x); verbs like *parol-i* ‘to speak’ < PAROL (sit) and *reĝ-i* ‘to be a king’ < REĜ (x); adverbs like *alt-e* ‘highly’ < ALT (qual) and *knab-e* ‘as a boy’ < KNAB (x).⁶ The remaining, closed, word classes and a limited number of adverbs consist of morphemes which are directly operable as words in the syntax. Examples are the temporal adverb *nun* ‘now’, the locational *tie* ‘there’, personal pronouns such as *mi* ‘I’, numerals like *unu* ‘one’, prepositions like *de* ‘of’, conjunctions like *se* ‘if’ and interjections like *ho* ‘hey’. These words are excluded from the present study, which is restricted to bound lexical morphemes as inputs to content words.

Esperanto is an agglutinating language, and word building (beyond the addition of a word class marker to a root as a mere act of syntactic preparation) involves derivation and compounding at a large scale. As to the scope of applicability of these two terms, there is no universal consensus in the referenced Esperanto literature. One reason for this can be traced back as far as Zamenhof (2004 [1887]). In the preface to this booklet Zamenhof writes: “For reasons of convenience, these prefixes and suffixes are given the meaning of autonomous words, and as such they were entered into the dictionary” and “I achieved a complete segmentation of concepts into autonomous words in such a way that the whole language consists of invariant words only, instead of words in their different grammatical forms” (Zamenhof 2004 [1887]: 9, 11).⁷ And even though, on these pages, Zamenhof continues to use the expressions ‘prefix’ and ‘suffix’, he states that the noun *frat-in-o* ‘sister’ “consists in reality of three words: *frat* (brother), *in* (woman, female), *o* (what is, exists)”. What is beyond dispute in Zamenhof’s morphology is the invariance of the elements used for word building. His terminology, however, is unfortunate in terms of modern linguistics. Although it may make sense to register the elements *frat*, *in* and *o* in these given forms in a dictionary, none of them is a word in any accepted definition of this term. On top of this, they are all known in the Esperanto literature (and in this study too) under three distinct denominations, i.e., ‘root’, ‘suffix’ and ‘ending’ respectively. The fact that not only these, but all other lexical and grammatical morphemes in the language are invariant components of an agglutinating structure, in no way justifies calling them all ‘words’. Admittedly, the noun marker *o* in *frat-in-o* hardly suffers from this confusion: Esperanto speakers will never use it in the same context as a content word or

⁵ Outside literal quotations from sources which may have adopted different styles, I use small caps to identify lexemes.

⁶ Whenever functional, explanatory hyphens are used inside complex words or lexemes.

⁷ From the Russian original. All translations are mine, unless otherwise stated.

even grammatical word, but only reflexively in expressions like ‘every noun ends in an *o*’.⁸ But, between *frat* and *in* the situation is different. Roots and affixes both carry a semantic charge, and for this reason some users may be tempted to copy the reference to both of them as ‘words’.

With respect to compounding, Zamenhof (2004 [1887]: 38–39) states in Rule 3 of the general part of his grammar: “Compound words are formed through a straightforward concatenation of words which are written together, but with their components separated by a fine stroke.”⁹ The most important word comes at the end. The grammatical endings are also considered as autonomous words, e.g., *vapor-ŝip-o* (‘steamship’) consists of *vapor* ‘steam’, *ŝip* ‘ship’ and *o* ‘ending of the noun.’ According to this rule, Zamenhof’s compounds are endocentric, and the internal head-modifier sequence is MH, i.e., from left to right. The head determines the semantic category of the compound and therefore its primary word class encoding. I will not dwell on the issue of Zamenhof’s terminology any longer than was needed here to show to what extent it may have contributed to differences in the interpretation of the language’s word building system. Compounding is addressed in a number of works on grammar following, among others, Saussure (1911, 1914, 1915, 1916) and Fruictier (1930), among which I will quote in the first place from the latest edition of Kalocsay & Waringhien (1985)¹⁰ and Wennergren (2005). All of these treat compounding as a grammatical process of combining *words*, and seek to apply the methodology and terminology of traditional European didactic grammar descriptions to linguistic features of Esperanto. None of them draws on a particular universal language theory, neither generative, nor functional or other.

2 Structure of the study

The study is structured as follows. In Section 3 I describe my approach to compounding in Esperanto as a lexical process. Section 4 makes ample use of textual examples of compounds retrieved from Kalocsay & Waringhien (1985) and Wennergren (2005). I decided for this approach to preserve a close link with the two books which are considered by many in the Esperanto community as their most authoritative grammatical references. Faced with a myriad of

⁸ ‘Reflexively’ in the sense of Lyons (1977: 5–6), i.e., in this expression, the ‘*o*’ is *mentioned*, not *used* as a noun marker.

⁹ In order to unify the annotation system in this paper, I replace Zamenhof’s fine vertical strokes by hyphens.

¹⁰ This work appeared in three editions under the title *Plena Gramatiko* ‘Complete Grammar’ and two revisions called *Plena Analiza Gramatiko* ‘Complete Analytical Grammar’. The References at the end of this paper contain the very first and very last edition from this series of five.

compounds that can be retrieved from the literature, many of them crystal-clear, others less transparent, it seemed to me efficient to stick to examples the reader may easily find back in said references. Section 5 describes the problem of epenthetic or bridging vowels between components of a compound, and Section 6 formulates the conclusions.

3 Compounding — the point of departure

Hengeveld & Mackenzie (2008) use the diamond symbol ‘◆’ for the head of a lexical property at the representational level (RL) of the grammar, i.e., a lexeme retrieved from the pool of primitives (fed by the lexicon) and inserted into a designated slot at the RL. A lexeme has a lexical definition, also referred to as its lexical content or meaning. The (abstract) lexeme is associated with a lexical morpheme, which I call ‘root’ in this study.

In Jansen (2016a), it is argued that derivation in Esperanto can be described efficiently if we assume it is a process which is hosted in the lexicon. A wide variety of (lexical) affixes is available for modifying base lexemes in order to expand the vocabulary of the language. Thus, the single root *ŜIP* can be modified by the suffix *ET* (diminutive) as *ŜIP-ET* ‘small ship’, by the suffix *AR* (collective) as *ŜIP-AR* ‘fleet’, or by the combination of the two as *ŜIP-ET-AR* ‘fleet of small ships’ or *ŜIP-AR-ET* ‘small fleet’. Conversely, it is possible to apply affixes to a combination of lexemes. If we take Zamenhof’s compound *VAPOR-ŜIP* from Section 1, derivation after compounding is illustrated by *VAPOR-ŜIP-AN* ‘crew member of a steamship’, in which *AN* (member, follower, inhabitant) can only apply to the finished compound. In a theoretically possible alternative approach, *AN* is applied to *ŜIP* first and yields *ŜIP-AN* ‘crew member’, in itself a perfectly legitimate derivation. Modification by *VAPOR* ‘steam’ would then reflect the meaning ‘ship’s crew member somehow related to steam’, which is nonsensical. The fact that in *VAPOR-ŜIP-AN* the couple *VAPOR-ŜIP* must have been completed before *AN* can be applied, suggests that the process of compounding, like the process of derivation, is hosted in the lexicon.

In terms of FDG, a compounding process in Esperanto puts two lexemes together in the linear sequence 1–2 from left to right and combines them into a new lexeme, expressing a single new concept. Compounds, in which one of the contributing elements can be assigned the head (H) role in the new lexeme, and the other its modifier (M), are called endocentric. For endocentricity between lexemes, I use the same symbol “:”, which is used in FDG for modification. Thus, Zamenhof’s *VAPORŜIP* is composed of the head *ŜIP* (Zamenhof’s “most important word comes at the end”) and the modifier *VAPOR*, i.e., *ŜIP : VAPOR*. The compounding process is then symbolized by (1):

$$(1) \quad \blacklozenge_2 : \blacklozenge_1 > \blacklozenge_3$$

In plain text: the indispensable H $\hat{\text{SIP}}$ is modified by the optional M VAPOR, restricting the set of ships to the subset of those which are propelled by a steam engine:

$$(2) \quad \hat{\text{SIP}} : \text{VAPOR} > \text{VAPOR}\hat{\text{SIP}}$$

If the meaning of the compound cannot be represented by a paraphrase derived from the apparent HM-distribution between the two components, i.e., cannot be described as the meaning of H restricted by M, then there is no hierarchical relationship between these components. The compound is exocentric, and exocentricity between lexemes I identify by the “+” sign between them, symbolizing the appositional nature or jointness of the two components. As examples I quote *soldat-kamarad-o* ‘soldier-comrade’ or ‘comrade in arms’ in (3) and *franc-german-aj (rilatoj)* ‘Franco-German (relations)’ in (4):

$$(3) \quad \blacklozenge_1 + \blacklozenge_2 > \blacklozenge_3$$

$$\text{SOLDAT} + \text{KAMARAD} > \text{SOLDATKAMARAD}$$

$$(4) \quad \blacklozenge_1 + \blacklozenge_2 > \blacklozenge_3$$

$$\text{FRANC} + \text{GERMAN} > \text{FRANCGERMAN}$$

The compound in (3) does not signify a specific kind of comrade, but a comrade who is a soldier at the same time. The meaning of (4) is ‘French (with the Germans) and German (with the French)’. In neither case is the first component restricted by the second, or the other way round.

Drawing on word-level examples retrieved from the reference literature, the formation of putative endocentric and exocentric compounds is described, including the way a lexically finished product is fed through the pool of lexical primitives into the grammar, where it is processed through the RL, encoded at the morphosyntactic level (ML) and at the phonological level (PL). At all three levels, the compound-internal source components play no further role, but until the articulation result is reached (represented in writing) hyphens will divide the compound into meaningful morphemes or syllables, whenever this is functional.

4 Reviewing compounds in Esperanto

In order to make sure that an overview of compounds in Esperanto is achieved which is as complete as possible, I will follow a simple procedure. I distinguish lexemes which on their own would designate one of three broad semantic

categories at the RL of the grammar: (i) Entities, (ii) Qualificational Dependents and (iii) Situational Dependents.¹¹ Entities cover not only First-Order Individuals (x), Second-Order States-of-Affairs (e) and Third-Order Abstracta (p), but may include Location (l), Time (t), Instrument (ins) and other independent properties in a typical Entity role or slot. The components $\blacklozenge_1$ and $\blacklozenge_2$ may come in any combination of each of these three broad categories. For a complete coverage the nine possible combinations will be addressed in Section 4.1 through 4.9.

4.1 Entity-Entity compounds

The first group I will discuss contains compounds whose second component on its own would designate an Entity, preceded by a component that would designate an Entity too. Compounds of this type are mostly determinative and endocentric. They are very frequent in Esperanto. The introductory noun *vaporŝip-o* ‘steamship’ from Zamenhof (2004 [1887]) is one of them. Other examples of Entity-Entity compounds are the nouns *vel-ŝip-o* ‘tall ship’ (VEL ‘sail’), *lum-radi-o* ‘light beam’, *hom-kor-o* ‘human heart’, *ŝton-pont-o* ‘stone bridge’, *paper-skatol-o* ‘paper box’, *lok-nom-o* ‘place name’, *hor-zon-o* ‘time zone’ and *krajon-strek-o* ‘pencil stroke’. In all cases, the right-hand component (head) is restricted by the left-hand component (modifier), i.e., the MH order is consistently LR. The restriction is relational in a very broad sense: it may be possessive, material, locative, destinative, etc., and can often be captured in an explanatory paraphrase including a dedicated preposition: *hom-kor-o* ‘heart of man’, *paper-skatol-o* ‘box out of paper’, *lok-nom-o* ‘name of a place’, etc. The analysis of these compounds follows (5):

- (5) a. $\blacklozenge_{2,x} : \blacklozenge_{1,x} > \blacklozenge_{3,x}$ ¹²
 b. $\hat{S}IP : VAPOR > VAPOR\hat{S}IP$
 c. RL ($x_3 : (f_3 : VAPOR\hat{S}IP (f_3)) (x_3)$) VAPOR $\hat{S}IP$
 d. ML ($N_3 : [(S_3 : vapor-\hat{s}ip (S_3)) (Infl_3 : o (Infl_3))] (N_3)$)¹³ *vapor-ŝip-o*
 e. PL ($PW_3 : /va-por'\hat{s}i-po/ (PW_3)$)¹⁴ /va-por'\hat{s}i-po/
 f. vaporŝipo ‘steamship’ *vaporŝipo*

¹¹ For the semantic categories which I consider relevant in Esperanto, see Jansen (2013) and the relevant remarks in Section 1 of this paper.

¹² $\blacklozenge_{i,x}$ means the i-th component of the compound, by itself designating an Individual (x) at the RL, etc.

¹³ I use N (and V, A and Adv) to denote *words*, without the need for specifying Nw for a noun word, etc. In Esperanto, N, V, A and Adv can only apply to morphosyntactic words and not to roots. The abbreviation ‘S’ for ‘stem’ is maintained, as it fits *stamo* ‘complex root’ in Esperanto, a synonym of *radikalo* (Koutny 2008: 35).

¹⁴ Because of the phonemic character of the Esperanto alphabet there is no immediate need for using phonemic symbols. At the PL, the normal letters of the alphabet are used, including hyphens to separate syllables, and the word accent sign is placed before the stressed syllable.

The way the compound lexeme VAPORŜIP enters the grammar simulates a situation in which an Esperanto speaker is asked the question “How do you say ‘steamship’ in Esperanto?”. The answer to this question is the reflexive *vaporŝipo*. In (5c), VAPORŜIP is inserted as the compound lexical head of the property f_3 , or lexeme, which designates the Individual x_3 at the representational level RL. At the morphosyntactic level ML in (5d), we find the primary word encoding or noun *vapor-ŝip-o*, which is the default encoding in this reflexive use of VAPORŜIP. At the phonological level PL in (5e), the phonological word is displayed, and in (5f) the written representation of the articulated word. The forms in the right-hand side column of (5c)–(5f) are the ones which I will use in the text outside the numbered examples.

The copulative *soldat-kamarad-o* of (3) is exocentric and is the result of $\blacklozenge_{1,x} + \blacklozenge_{2,x} > \blacklozenge_{3,x}$, and so is, e.g., *kverk-arb-o* ‘oak tree’ $< \text{KVERK} + \text{ARB}$. One may argue over the latter compound whether it is indeed exocentric (as suggested by Kalocsay & Waringhien 1985) or endocentric, i.e., restricting the set of trees to the subset of oaks, but the issue seems to me theoretical rather than having any practical value or consequences.

4.2 Qualificational-Entity compounds

These compounds, less common than the ones above, involve a first component which on its own would designate a Qualificational dependent (qual) at the RL, with primary encoding as an adjective in the syntax. The consecutive insertion of such a pair of lexemes from the lexicon through the pool of primitives would ultimately generate in the grammar a noun phrase (NP) with a head noun and an adjective. Examples of this are *varm-a veter-o* ‘warm weather’ and *bel-a tag-o* ‘nice day’. Compounding in the lexicon, however, prepares for a simultaneous insertion of the two as a single block. Kalocsay & Waringhien (1985) illustrate this by the noun *varm-energi-o*, not meaning ‘hot energy’ (which would obviously be the meaning of the phrase *varm-a energi-o*), but ‘heat energy’, i.e., ‘energy (generated) out of heat’. Another example of a compound of this not very productive type is *bel-kult-o* ‘beauty cult’. Yet, Esperanto lexemes being contentive, nothing speaks against using Qualificationals like VARM and BEL as modifiers inside an endocentric determinative compound, at which level we are concerned with conceptual compounding. With ‘heat energy’ and ‘beauty cult’ as the universally accepted meanings of these compounds, Kalocsay & Waringhien use the paraphrases *energio de varm-o* ‘energy from heat’ and *kulto de bel-o* ‘beauty cult’ for explanatory purposes, in both cases with nominal applications of the Qualificational. It is arguable that *energio de varm-ec-o* and *kulto de bel-ec-o* might be more appropriate, since these explicitize the derived Abstracta VARM-EC (p) and BEL-EC (p), but the expression or not of such a specificity seems to me the prerogative of a phrase in its grammatical and

extragrammatical context (e.g., are we discussing the heat energy of a gas or heat energy as a concept?), and not necessarily required in a compound. Nothing, by the way, prevents us from coining the term *varm-ec-energio*, if required. The analysis of *varm-energi-o* is given in (6):

- (6)
- | | | |
|----|---|----------------------|
| a. | $\blacklozenge_{2,p} : \blacklozenge_{1,qual} > \blacklozenge_{3,p}$ | |
| b. | ENERGI : VARM > VARMENERGI | |
| c. | RL (p_3 : (f_3 : VARMENERGI (f_3)) (p_3)) | VARMENERGI |
| d. | ML (N_3 : [$(S_3$: varm-energi (S_3)) (Infl ₃ : o (Infl ₃))] (N_3)) | <i>varm-energi-o</i> |
| e. | PL (PW_3 : /varm-e-ner'gi-o/ (PW_3)) | /varm-e-ner'gi-o/ |
| f. | varmenergio ‘heat energy’ | <i>varmenergio</i> |

There are other determinative compounds built on a Qualificational-Entity sequence which cannot be accounted for the same way as *varm-energi-o*. As an example, I take the complex noun *alt-forn-o* ‘blast furnace’ (Wennergren 2005), built on ALT ‘high, tall’ and FORN ‘furnace’. An *alt-forn-o* is not just a furnace that happens to be high, but a furnace that belongs to a category of high-rise furnaces, whose height is by no means an accidental property, but is required by the nature of the chemical processes that take place inside. In fact, there may be relatively high and also lower blast furnaces, whereas a high furnace can’t be low at the same time. *Alt-forn-o* ‘blast furnace’ (German *Hoch-ofen*) and examples like *liber-temp-o* ‘spare time’ (German *Frei-zeit*) and *grand-urb-o* ‘metropolis’ (German *Groß-stadt*) seem to have entered Esperanto as calques from German, as suggested by the translations I added. Of definitely non-German, but Swedish origin are the concepts of a *long-dom-o* or ‘long house’ in old Nordic architecture, and the *long-ŝip-o* ‘long ship’ of the Vikings, found in Bertil Nilsson’s high-quality Esperanto translation of the novel *Röde Orm* [*Orm the Red*] (Bengtsson 1989).

The number of compounds of this kind is limited and is given by those few candidates in a number of source languages which are deemed worth taking over in Esperanto. Thus, the German *Weißwein* ‘white wine’ has not made its appearance in Esperanto in the form of **blank-vin-o* (the correct expression is the phrase *blank-a vin-o*) and *Rotkohl* ‘red cabbage’ has not been borrowed as **ruĝ-brasik-o* (it is *ruĝ-a brasik-o*). Structures like **blank-vin-o*, **ruĝ-brasik-o*, **bon-libr-o* ‘good book’ or **bel-veter-o* ‘nice weather’ are rejected by the speakers. The structure of the accepted compounds clearly reveals their etymology as endocentric compounds loaned to Esperanto as single blocks, not as separate components. For this reason, I propose to call them ‘loan compounds’.

There is yet another family of Qualificational-Entity compounds. This is exemplified by LONG-NAZ ‘long-nosed’, which differs fundamentally from the

VARM-ENERGI type and from the small class of the GRAND-URB type loan compounds. They manifest themselves as adjectives, despite the fact that their second component is an Entity.¹⁵ Compounds of this type are attested starting from the earliest documents, and they feature in the more modern grammar descriptions of Kalocsay & Waringhien (1985), Wennergren (2005) and others, with textual examples like *facil-anim-a* ‘light-hearted’ and *blank-ĉapel-a* ‘with a white hat’. Based on the function they clearly display both in translation and in Esperanto, I will refer to them as possessive compounds. Because of their peculiar nature I will subject them to a more detailed approach first, before proposing their classification.

Supposing LONG-NAZ is endocentric with NAZ as its regular (right-hand) head and LONG as its modifier, then LONG-NAZ would have to be encoded primarily as the noun *long-naz-o*. This, however, is not the case: the existence of *long-naz-o* is not attested. On the other hand, this does not mean that *long-naz-o* is an impossible word. If, for whatever reason, *alt-forn-o* and *long-ŝip-o* exist, there is no reason why *long-naz-o* ‘long-nose’ should be declared inadmissible on principle. The only thing we can say at this moment is that the community has not yet found a reason to justify its introduction. If the theoretical grammaticality of *long-naz-o* is beyond doubt, what about *long-naz-a*? This is precisely the form that is encountered in the language, i.e., in a modifier slot. It is in this adjectival application that we are confronted with an old issue in the Esperanto grammar, which was already highlighted by the sample word *teatr-a* in Rule 15 of the grammar section of Zamenhof (1963 [1905])¹⁶, presented there as the equivalent of ‘theatrical’. In Esperanto’s contentive parts-of-speech system (Jansen 2013), the use of an Individual like TEATR in a modifier slot, encoded in the morphosyntax as *teatr-a*, just means ‘of the theater, in the theater’ or something similar, or ‘theater’ as the modifier in a compound like ‘theater play’. Sometimes, ‘theatrical’ can be used as a synonym, but if ‘theatrical’ is taken with the specific meaning ‘marked by exaggerated self-display and unnatural behavior; affectedly dramatic’, then *teatr-a* falls short, because it lacks the material content that would be required in this latter definition. A dedicated derivational tool could be expected, but is missing in Esperanto. In its place, the suffix *ec* is often used in similar circumstances (Jansen 2020). Formally speaking, *long-naz-a* as an endocentric compound can be used in an expression like *long-naz-a form-o* ‘the shape of the long-nose’, but not in the phrase *long-naz-a vir-o* ‘long-nosed man’, because, roughly speaking, this would turn the possessor into a possessed, and vice versa. The question now

¹⁵ This section was first presented for discussion and comments at the Buchanan Masterclass at the University of Liverpool, 2016, see Jansen (2016b).

¹⁶ It is *teatra* in four of the reference languages; in French, the grammar section has *tragedia* ‘tragic’.

is, why *long-naz-a* is used in the community, with the latter meaning ‘having a long nose’.

Since ‘long-nosed’ as the meaning of *long-naz-a* is observationally correct, we should ask ourselves why this is the case. It requires the language user to rely on the existence of a hypothetical *long-naz-o* (of which one wouldn’t really know how this relates to *long-a naz-o*) and to accept that *long-naz-a* as an Entity in a modifier slot is to be interpreted possessive instead of relational. The answer I propose is that this is first of all due to the mother tongue influence, or pressure of national languages in general, which Esperanto speakers, who are L2 users by definition, are constantly exposed to. These may provide attractive model solutions of Qualificational-Entity, or adjective-noun, compounds that are worth calquing. Among the close to one hundred compounds and incorporations that Neves (2011) quotes for the period 1887–1888, only seven are of this particular type, six from Zamenhof (1888): *ali-land-a* (R *ino-stran-nyj*) ‘of another country, foreign’, *ali-naci-a* (R *ino-rod-nyj*) ‘of another homeland, foreign’, *bon-gust-a* (G *wohl-schmeck-end*) ‘good tasting’, *bon-intenc-e* (G *wohl ge-mein-t*) ‘well-intended’, *bon-son-a* (G *wohl-kling-end*) ‘good-sounding’, *propr-a-man-e* (G *eigen-händ-ig*) ‘with one’s own hands’; and one from the first textbook of Esperanto for Germans (Einstein 1888): *sam-e-temp-e* (G *gleich-zeit-ig*) ‘at the same time’. Five out of seven closely follow a German model, the other two a Russian equivalent, identified by G and R respectively. Their common structure is ‘adjectival/adverbial–nominal, followed by an adjectivizing suffix or participle ending’. However, the final *a* in Esperanto *long-naz-a* and the final *a*’s and *e*’s in the seven attestations above may be *regarded* by speakers as equivalents of an adjectivizing suffix, they are strictly syntactic markers only. The tempting misinterpretation is the indirect consequence of the pressure of the predominantly Central European environment in which Esperanto matured, and of the conscious or subconscious uncertainty among its speakers as to the nature of the lexical items in the language (contentive or to some extent syntactically specialized?) and the precise role of the word class markers.

Could the combination LONG-NAZ be recognized as an exocentric compound designating an animate Individual characterized by the qualification of having a long nose? Such a procedure, often including a special stress pattern, is found in many different languages (think of ‘redneck’ and ‘paleface’ in English), but an Esperanto speaker would not accept this solution and spontaneously insert the appropriate suffix *ul* in *long-naz-ul-o* ‘long-nosed person’ (without realizing that this does not make *long-naz-a* any better).¹⁷ The

¹⁷ The lexical suffix *-ul* derives a complex root which expresses the animate individual that has the characteristics of the root it is suffixed to, e.g., AVAR ‘greedy’ > AVAR-UL ‘miser’. (Jansen 2016a).

circumstance that LONG-NAZ-UL is the unquestioned form denoting a person with a particular qualification, defines LONG-NAZ as this qualification. This points to the solution we are looking for. The compound LONG-NAZ is exocentric, but semantically categorized as a Qualificational and encoded as an adjective in the syntax. This may be in contrast with certain nominal solutions in modern English, which are mentioned above, or with similar solutions in contemporary German, but they do conform to such adjectival structures like the English ‘long-nosed’ itself, German *schwarz-äug-ig* ‘with black eyes’ (Esperanto *nigr-okul-a*), and many more, following the primordial examples registered in 1887–1888. In sum, the exocentric analysis of LONG-NAZ is as symbolized in (7):

- (7) a. $\blacklozenge_{1,qual} + \blacklozenge_{2,x} > \blacklozenge_{3,qual}$
 b. LONG + NAZ > LONGNAZ¹⁸

Because endocentric compounds of the GRAND-URB type have a minimum level of expected productivity, and exocentric compounds of the LONG-NAZ type can be created at will, newly coined compounds of the Qualificational-Entity type can be of either type, and ambiguity is not excluded. I will come back to this issue in the conclusions of Section 4.10.

4.3 *Situational-Entity compounds*

An example of this family is the noun *lern-procez-o* built on PROCEZ (p) ‘process’ preceded by LERN (sit) ‘learn’. The contentiveness of Esperanto lexemes allows a Situational like LERN to be used as a modifier inside an endocentric determinative compound. The meaning of the compound is ‘learning process’, paraphrasable as *procez-o de lern-o*. The situation is reminiscent of what was discussed in the context of (6): had the speaker targeted the production of a phrase, *procez-o de lern-ad-o* might have been preferred, with the lexically derived State-of-Affairs LERN-AD in the semantics and the noun *lern-ad-o* as its primary encoding in the syntax. In the compound, however, LERN as a modifying component is appropriate. The elaboration in (8) is similar to what we found in (6), and only the first two lines will be detailed:

- (8) a. $\blacklozenge_{2,p} : \blacklozenge_{1,sit} > \blacklozenge_{3,p}$
 b. PROCEZ : LERN > LERNPROCEZ

¹⁸ From now on I will not repeat the lines c-f which I showed in (5) and (6) because the grammatical insertion follows the same pattern in this and other cases to come.

Some other examples of compounds of this type are the nouns *ir-manier-o* ‘way of going’, *flug-boat-o* ‘boat for flying = flying boat’, *skrib-maŝin-o* ‘machine for writing = type-writer’.

4.4 Entity-Qualificational compounds

In this Section I will discuss compounds whose second component, when entered into the grammar on its own, would designate a Qualificational dependent, earmarked for primary encoding as an adjective. The accompanying first component is an Entity. It appears that the qualification expressed by the second component is further modified by the first, and that the sequence first-second reflects the modifier-head sequence in an endocentric determinative compound. An example of a compound of this type is *HERB-RIC* ‘rich in grass’ or ‘abundant in grass’ (Kalocsay & Waringhien 1985), in which the modification of the head *RIC* ‘rich’ restricts the lushness of the vegetation to the abundant presence of *HERB* ‘grass’. The lexical process is illustrated by (9):

- (9) a. $\blacklozenge_{2,qual} : \blacklozenge_{1,x} > \blacklozenge_{3,qual}$
 b. *RIC* : *HERB* > *HERBRIC*

Other examples of compounds of this type are the adjectives *di-simil-a* ‘god-like’, *sang-avid-a* ‘blood-thirsty’ and *sang-ruĝ-a* ‘bloodred’.

4.5 Qualificational-Qualificational compounds

I distinguish two subsets in this category. The first one can be exemplified by the adjective *hel-ruĝ-a* ‘light red’, built on *HEL* ‘light, clear’ and *RUĜ* ‘red’. It is sometimes argued that the phrase *hel-e ruĝ-a*, with the adverbial modifier *hel-e* identifies a different shade of light red than the compound *hel-ruĝ-a*, but this seems to me completely open to discussion. The Esperanto community has no unified color experience of its own, and with inputs from so many different cultures with varying interpretations of the color spectrum in the nomenclature each of them uses, it seems to me pedantic to prescribe a difference which may exist for some and not for others. For the time being I propose to classify these as endocentric compounds and lexical alternatives to a grammatical (phrasal) expression.

A different approach is required for the second subset. Compounds of this type are copulative, comparable to the subset discussed in Section 4.1. The example which I present here is the adjective *surd-mut-a* ‘deaf-and-dumb’ (adapted from Kalocsay & Waringhien 1985), built on *SURD* ‘deaf’ and *MUT* ‘dumb, mute’, jointly constituting the head of the resulting lexical property, i.e., the exocentric compound lexeme:

- (10) a. $\blacklozenge_{1,qual} + \blacklozenge_{2,qual} > \blacklozenge_{3,qual}$
 b. SURD + MUT > SURDMUT

The directionality in *surd-mut-a* is conventional and follows national language patterns like ‘deaf-and-dumb’, French *sourd-muet*, German *taubstumm*.

In some cases, such exocentric compounds are used for a specific expressive purpose. In *plen-plen-a* ‘overcrowded’ the reduplication of the root PLEN ‘full’ intensifies the meaning or effect that it has by itself, and is more or less synonymous with the augmentative derivation in PLEN-EG, encoded as the adjective *plen-eg-a* (cfr. Spanish *llen-isim-o*).

4.6 *Situational-Qualificational compounds*

An example of a compound of this type is the adjective *pres-pret-a* ‘ready for printing’ (Kalocsay & Waringhien 1985), which is built on the Situational PRES ‘print’ followed by the Qualificational PRET ‘ready’. There is a clear parallel with the discussion about *lern-procez-o* in Section 4.3. If the speaker aims at the production of a phrase, *pret-a por pres-ad-o* with the preposition *por* ‘for’ governing the noun *pres-ad-o* built on the State-of-Affairs PRES-AD, may be the preferred formulation. In the compound, however, PRES as a modifying component is appropriate. The analysis of *pres-pret-a* starts with:

- (11) a. $\blacklozenge_{2,qual} : \blacklozenge_{1,sit} > \blacklozenge_{3,qual}$
 b. PRET : PRES > PRESRET

Other examples of a compound of this type are *flug-pret-a* ‘ready to fly’, *parol-kapabl-a* ‘able to speak’, *atest-matur-a* ‘mature to testify’.

4.7 *Entity-Situational compounds*

These are compounds whose second component would designate a Situational dependent, when entered into the grammar on its own, preceded by a component that would designate an Entity. It appears that the second component is indeed modified by the first, and that the sequence first-second reflects the internal modifier-head sequence in an endocentric determinative compound. Often the modification of the head is instrumental. An example is *šip-vetur-i* ‘to travel by ship’ < ŠIP-VETUR, in which the head VETUR ‘go, drive, sail’ is modified by ŠIP ‘ship’, restricting the travelling to that by ship (Waringhien 2002):

- (12) a. $\blacklozenge_{sit2} : \blacklozenge_{x1} > \blacklozenge_{sit3}$
 b. VETUR : ŠIP > ŠIPVETUR

Other examples of compound verbs of this type are *pied-ir-i* ‘to foot-go = to walk’, *milit-ir-i* ‘to war-go = to conduct a campaign’, *ink-skrib-i* ‘to ink-write = to write with ink’, *muzel-fos-i* ‘to muzzle-dig = to burrow’ and *martel-bat-i* ‘to hammer-beat = to beat with a hammer’.

4.8 *Qualificational-Situational compounds*

Another group of compounds within the Situational-final family is introduced by Kalocsay & Waringhien (1985) through the verb *laŭt-leg-i* ‘to read (out)’, with its specific connotation ‘to read a story to children’ or ‘to read out a speech or a proclamation in public’, a clear calque from Russian. The special meaning of this compound compared with the phrase *laŭt-e leg-i* ‘to read loudly’ is confirmed by the fact that *laŭt-leg-i*, in contrast with the phrase *laŭt-e leg-i*, can be further modified by *laŭt-e* (‘to read out loudly’) or *mal-laŭt-e* (‘to read out softly’). This shows a perfect parallel with the distinction made between *alt-forn-o* and *alt-a forn-o* in Section 4.2. Therefore, I propose again that combinations like LAŬT-LEG, whose verbal applications are registered in the literature or in authoritative dictionaries of Esperanto with a special meaning other than that of a seemingly equivalent phrase be considered as endocentric loan compounds.

There are cases in which there is no clearly identifiable difference in meaning between a compound of the type described above and the phrase composed of the same lexical items, e.g., *bon-fart-i* ‘to do well, to do fine’ and *bon-e fart-i* (Kalocsay & Waringhien 1985). The reason for this is probably that the verb *fart-i* ‘to be doing (healthwise)’ never occurs without an adverbial modifier. The indispensability of joining *bon* to *fart* at the phrasal level means that the link is as strong as a word-internal link between the same lexical components. Another case is *bon-odor-i* ‘to smell good’ which, in practice, is equivalent to *bone odori*. As in the case of *hel-ruĝ-a* discussed in Section 4.5, I consider these as endocentric determinatives which do not require a separate treatment and can be dealt with as lexical alternatives to a grammatical (phrasal) expression.

4.9 *Situational-Situational compounds*

Just like the subset in Section 4.1 (*kamarad-soldat-o*) and the compounds of Section 4.5 (*surd-mut-a*), compounds of two lexemes which both designate a Situational dependent, produce an exocentric copulative compound. Examples from Kalocsay & Waringhien (1985) are *paf-murd-i* ‘to murder by shooting’ (litt. ‘to shoot-kill’) < PAF-MURD and *pet-vetur-i* ‘to hitch-hike’ (litt. ‘to request-drive’) < PET-VETUR. Attempts by Kalocsay & Waringhien to classify these compounds as endocentric (without using this term) fail: *pet-vetur-i* is not the same as *pet-e veturi*, which is an approximative paraphrase for illustration

purposes only. In the compound the link between the two lexical components is much stronger, less occasional, less contingent than in the phrase. An internal HM assignment is not possible and both Situationals together constitute the new (Situational) head of the resulting lexical property or compound lexeme:

- (13) a. $\blacklozenge_{1,\text{sit}} + \blacklozenge_{2,\text{sit}} > \blacklozenge_{3,\text{sit}}$
 b. PET + VETUR > PETVETUR

4.10 Summary of compounds with two lexical components

Sections 4.1–4.9 can be summarized in the following two tables. The majority of compounds in Esperanto are found to be of the endocentric determinative type (Table 1). A large-scale corpus analysis (which has not been performed for this study) will no doubt demonstrate that the type majority is reflected by an overwhelming frequency of occurrence of different tokens of compound nouns, adjectives, verbs and adverbs of this type. In retrospect we may state that Zamenhof's (2004 [1887]) single rule on compounding was destined to cover most of the compounds that were to be introduced in the language by its users. Endocentricity allows the hearer to decode the meaning of a compound used by the speaker in a straightforward process of restricting the right-hand head by the left-hand modifier, and constitutes a transparent procedure.

Table 1: Prototypes of endocentric determinative compounds in Esperanto.

M	H	Primary word	Meaning
VAPOR	ŜIP	<i>vapor-ŝip-o</i>	steamship
VARM	ENERGI	<i>varm-energi-o</i>	heat energy
ALT	FORN	<i>alt-forn-o</i> ¹⁹	blast furnace
LERN	PROCEZ	<i>lern-procez-o</i>	learning process
HERB	RIĈ	<i>herb-riĉ-a</i>	rich in grass
HEL	RUĜ	<i>hel-ruĝ-a</i>	light red
PRES	PRET	<i>pres-pret-a</i>	ready for printing
ŜIP	VETUR	<i>ŝip-vetur-i</i>	to travel by ship
LAŬT	LEG	<i>laŭt-leg-i</i> ¹⁹	to read out
BON	ODOR	<i>bon-odor-i</i>	to smell good

The rows in Table 2 give prototypes of the different forms of exocentricity that can be found in Esperanto.

¹⁹ Member of a small and hardly productive class of loan compounds.

Table 2: Prototypes of exocentric compounds in Esperanto.

M	H	Primary word	Meaning
SOLDAT	KAMARAD	<i>soldat-kamarad-o</i>	comrade in arms
SURD	MUT	<i>surd-mut-a</i>	deaf and dumb
PET	VETUR	<i>pet-vetur-i</i>	to hitch-hike
LONG	NAZ	<i>long-naz-a</i>	long-nosed

Instantiations like *soldat-kamarad-o*, *surd-mut-a* and *pet-vetur-i* show that this form of exocentricity occurs in the event of a combination of two components which, when inserted in the grammar on their own, designate one and the same semantic category: Entity-Entity, Qualificational-Qualificational or Situational-Situational. Wennergren (2005) correctly identifies a grey zone between appositional structures and compounds. The meaning of such compounds is given by the simple sum of two meanings plus, maximally, a conventionalized connotation (but not in, e.g., *surd-mut-a*). Eventually, it is a combination of world knowledge and convention which will guide our choice, but in didactic grammars these exocentric expressions can be treated as pretty self-explanatory, i.e., as the algebraic sum of the two component meanings.

The exocentricity of the bottom row is different. It is a clear violation of Zamenhof's Rule 3, which specifies that compounds are MH structures. These compounds are the only complex and purely lexical items whose semantic categorization is not defined by the right-hand component because it does not constitute the head. LONG-NAZ being a Qualificational, one might be tempted to consider *long-naz-o* as its the nominal application, just like *bel-o* 'that which is nice' is the nominalization of the Qualificational BEL 'nice', in contrast with the Abstractum BEL-EC. However, this clashes with an endocentric interpretation of *long-naz-o*, which is still possible in analogy with the *alt-forn* or *grand-urb* type. This is a consequence of the exceptional status of exocentricity in compounding. The alternative *long-naz-ec-o* offers itself, and is actually used, with a loss of distinctive capability between a nominalized Qualificational and an Abstractum. Conversely, one may say that something similar can be expected with *grand-urb-eco*, if a speaker is not sure about the status of GRAND-URB: with a primary exocentric *grand-urb-a* 'metropolitan' or endocentric *grand-urb-o*.

5 Epenthetic or bridging vowels

"FDG is a theory about grammar, but one that tries to reflect psycholinguistic evidence in its basic architecture" (Hengeveld & Mackenzie 2008: 2). The chosen architecture represents a top-down process, which starts with intentions and ends with the articulation, a model which is suggested by psycholinguistic

research. This approach also means that the plausibility of hypotheses set up by language researchers should not only meet certain model-internal coherence requirements, but also that consistency with psycholinguistic data can be demonstrated. Let's see what this teaches us about an issue that is often glossed over in Esperanto compounding, the choice of a bridging vowel (when needed).

Given the circumstance that the formulation and shaping of a linguistic expression from its inception up to its final articulation is often completed within a timespan of seconds or less, it is assumed in FDG that a speaker cannot afford wasting any time or energy by building up certain provisional structures with intermediate results, which are kept active only as long as they are needed, to be retrieved and replaced or suppressed as the speaker proceeds. The non-availability of time is one factor. Another one is of a methodological nature: provisional structures may come in handy, but their very existence cannot be proven since they disappear by virtue of their provisionality and leave no trace behind (other than the desired result which was their justification, thus leading to a circular argumentation). If allowance is made for the existence of provisional structures, there is no limit to what a researcher may come up with in order to support an idea. It is for this reason that in FDG the approach is taken that linguistic phenomena should be investigated and explained without recourse to provisional structures and without making U-turns in the grammar or between the grammar and the lexicon, e.g., for the purpose of retrieving and correcting intermediate results. This approach partly contradicts the spirit of a famous set of principles in the traditional Esperanto morphology, the so-called Principle (or, rather, Pair of Principles) of Necessary and Sufficient Conditions, proposed by the Swiss mathematician René de Saussure and approved by the Esperanto Academy in 1913 (Kalocsay & Waringhien 1985: 387):

- a. "Principle of Necessary Conditions: in the process of word building one must put together all roots, suffixes, prefixes and endings which are necessary to clearly and fully evoke the idea reflected by this word."
- b. "Principle of Sufficient Conditions: should the idea thus expressed be properly and unmistakably understandable from the context, and without one of the suffixes used, this suffix can be eliminated as a useless and superfluous element."

The key to our discussion about provisional structures can be found in the last part of the Principle of Sufficient Conditions: "this suffix can be eliminated as a useless and superfluous element", which would mean that in a spontaneous process of speech production allowance is made for the creation of provisional

structures. Following in De Saussure's footsteps, other grammarians²⁰ further elaborated a word building theory for Esperanto, based on syntactically specialized lexemes and, for what concerns compounding, the combination of syntactic words.²¹ In other words, a compound word like *man-tuk-o* 'towel' is built by joining the word *man-o* 'hand' to the word *tuk-o* 'cloth', and suppressing the *o* at the boundary between the two because the resulting *man-tuk-o* is well pronounceable. *Mantuko* is explained as a *tuko por (la) mano(j)* 'cloth for (the) hand(s)', i.e., it is one of a subset of cloths used for drying one's hands. The suppression of the *o* in the middle doesn't always work, e.g., in the compound *tabl-o-tuk-o* 'table cloth', its alternative **tabltuko* being difficult to pronounce. Copying from *tuko por tablo* 'cloth for (a) table', the *o* is selected as the bridging vowel to be inserted. My alternative analysis of the compound *tablotuko* is given in (14) below:

- (14) a. $\blacklozenge_{2,x} : \blacklozenge_{1,x} > \blacklozenge_{3,x}$
 b. TUK : TABL > TABLTUK
 c. RL (x_3 : (f_3 : TABLTUK (f_3)) (x_3)) TABLTUK
 d. ML (N_3 : [$(S_3$: tabl-tuk (S_3)) ($Infl_3$: o ($Infl_3$))] (N_3)) tabl-tuk-o
 e. PL (PW_3 : /ta-blo'tu-ko/ (PW_3)) /ta-blo'tu-ko/
 f. tablotuko 'table cloth' *tablotuko*

Although the bridging vowel issue is strictly related to the syllabification of a morphemic structure in order to make it pronounceable (*tabl-tuk-o* is morphologically immaculate and perfectly understandable in a written text), the traditional word theory in Esperanto forces us to seek a morphosyntactic justification for the vowel to be inserted, and to look for some equivalent phrase structure that should suggest the appropriate choice.²² *Tuko por tablo* ‘cloth for (a) table’ is such a phrase. Others are *vitro-domo* ‘glass house’ < *domo el vitro* ‘house from glass’, *dormo-kambro* ‘bedroom’ < *kambro por dormo* ‘room for sleeping’, and *nigra-blanka* ‘black-and-white’ < *nigra kaj blanka* ‘black and white’. But in the case of ‘textbook’, which is *lernolibro* in Esperanto, we encounter the same *o* in the middle, although there is no rule that says that the

²⁰ To mention just a few: Fruictier (1930) and Kalocsay & Waringhien in the five editions between 1935 and 1985.

²¹ See, e.g., Kalocsay & Waringhien (1985: 395): *La gramatikaj finaĵoj inter la elementoj estas forigitaj, se ilian konservon ne postulas la prononceblo aŭ aliaj konsideroj* ‘The grammatical endings between the elements [WJ: of complex words] are removed, unless their pronounceability or other considerations require them to be maintained’, suggesting that the elements used in word building are initially provided with endings (they are words), which they may or may not lose in the process.

²² See Oostendorp (1999) for syllabification in Esperanto. See Jansen (2015, 2016b) for earlier presentations of the bridging vowel issue.

phrase *libro por lern-o* is better suited than *libro por lern-i*. However, the latter phrase would have produced the uncommon **lern-i-libr-o*. Since most complex noun modifiers consist of a noun or NP headed by a preposition, *por lern-i* is avoided as a model in favor of *por lern-o*. De Saussure's argumentation goes further and proposes to interpret the phrase *libro por lern-o* as short for *libro por lern-ad-o*, from which the suffix *ad* is suppressed as being superfluous for a correct understanding of what is meant (see Section 4.3). This way, the loop inside the grammar to justify the bridging *o* is even doubled: first of all, a syntactic phrase structure is taken not as an explanatory paraphrase, but as the base on which the compound is built, and, secondly, a provisional structure is set up, including a conveniently selected suffix, to be suppressed in the end product. Needless to say, a much more appropriate analysis of *lernolibro* starts with the lexicon-hosted step (15),

(15) LIBR : LERN > LERNLIBR

which, however, does not resolve the bridging vowel issue.

There is more. The bridging *i* that was deemed undesirable in **lernilibro* does appear in compounds like *pag-i-pova* 'able to pay (in hard currency)', for which the model phrase *pov-ant-a pag-i* 'being able to pay' with the subsequent suppression of the present participle ending *ant* is obviously given priority over, e.g., *pov-a je pag-o*, which would be reminiscent of *čio-pova* and the phrase *pova je čio* 'all-mighty'.²³ In the compound *art-e-farita* (litt. 'art-made') 'artificial' we find a bridging *e*. We may wonder what then the difference is between the adjectival phrase *art-e farita* 'artificially made' with ART (x) and obligatory *e*-marking in the adverbial slot modifying *farita* 'made', and the one-word compound adjective *art-e-farita* 'artificial', and why *art-o-farita* with *farita per art-o* 'made by art' is rejected (if it is the formal shape of a phrase that is supposed to determine the choice of a bridging vowel). The traditional theory forces us, at the point of syllabification of the compound, to revert to the morphosyntax, to hypothesize equivalent phrases, to make choices between alternative candidates based on convention and analogy rather than anything else, and to make ample use of provisional structures. The very complexity of these processes should be a warning signal that there may be something seriously wrong with the underlying argumentation.

In the introduction to this study, and in particular in Section 3, I argued that the process of compounding is hosted in the lexicon and involves roots

²³ The final *o* in *čio* 'everything' and other correlatives from this series is indicative of the nominal slot in the syntax they occupy, but this *o* is not a formal ending and cannot be decoupled.

instead of words. Any combination of roots underlies a semantic relation of endocentricity or exocentricity, and represents a new concept. The more detailed nature of this semantic relationship can be illustrated by a phrase, but the phrase cannot be used as a substitute for the compound. A *mantuko* ‘towel’ can be described as a *tuko por la manoj* ‘piece of cloth for the hands’, but the internal bond in the lexical product MAN-TUK is much stronger than that of the grammatical construct *tuko por la manoj*, and, clearly, not every piece of cloth for the hands is a towel. Since compounding is a lexical process, there can’t be any internal inflection to be disposed of. There simply is no inflection, neither internal nor, before reaching the morphosyntactic level of the grammar, at the outer compound boundary, i.e., there is no *a*, *e*, *i*, *o* or *u* to be disposed of. What may be required at the point of syllabification is a way to overcome an occasional bad pronounceability of the compound due to an unfortunate consonant cluster at the internal root-root boundary.²⁴ To this end, languages often use bridging vowels, and so does Esperanto, as the tradition shows. Any vowel, selected on the basis of whatever criteria, may do the job, but unfortunately each of the five available vowels in Esperanto has a well-defined grammatical function, and this, in turn, makes it tempting to seek a syntactic justification for the selection of this mere pronunciation aid through a paraphrase.

In ‘received Esperanto’ there is no solution to this problem. We would need a syntax-free vowel, which we do not have. If we could make use of the *y* sign, which has semi- or full vowel value in many languages, and pronounce it, e.g., as /ə/, we would have **tablytuko*, **lernylibro*, **pagypova*, **artyfarita* and others, but this is sheer fantasy.

For the time being, we have no other choice than to accept the bridging vowels we are used to and which seem to be based on convention rather than anything else. Additional research may, however, reveal that the choice of an epenthetic vowel may be governed by a (set of) phonological rule(s) based on the nature of the consonant cluster at the root boundary and possibly other phonological factors.

6 Conclusions

The traditional Esperanto word building theory was sanctioned by the Esperanto Academy (*Aktoj 1963–1967*). In contrast with these proceedings the here presented description, which is based on root compounding, is believed to be a

²⁴ Such clashes cannot happen in derivation because derivational affixes are limited in number and tailor-made in order to avoid such occurrences. In compounding, we are faced with a theoretically unlimited number of combinatorial possibilities.

lot simpler. It clearly separates semantic from morphosyntactic aspects and specifies the grammatical characterization of roots as strictly semantic. It is claimed to be more credible in terms of representativity of the mental processes that underly compounding. It also does justice to the L2 role of Esperanto in its permanent interaction with L1's as model or feeder languages. The followed approach also lays bare a few areas (e.g., an improper derivational analysis of exocentricity, the vowel choice in epenthesis) which deserve more attention than they normally get.

It should be stressed that the findings in this paper do not affect the observationally accepted meaning of any compound quoted or retrievable in spoken or written received Esperanto. A new analysis, based on FDG, is proposed, which is claimed to have a greater explanatory value than any other current description of compounds in Esperanto.

References

- Aktoj de la Akademio 1963–1967* [*Proceedings of the Academy 1963–1967*]. 2007. Rotterdam: Akademio de Esperanto. Second edition.
- Bengtsson, Frans G. 1989. *Orm la Ruĝa* [*Orm the Red*]. Malmö: Eldona Societo Esperanto. Translated by Bertil Nilsson.
- Einstein, Leopold. 1888. *Plena lernolibro por Germanoj* [*Complete Textbook for Germans*]. Nürnberg: J.A. Stein.
- Fruictier, Paul. 1930. *Kompleta Gramatiko kaj Vortfarado de Esperanto* [*Complete Grammar and Word-building of Esperanto*]. Jaslo (Poland): Esperantista Voĉo.
- Hengeveld, Kees & Lachlan Mackenzie. 2008. *Functional Discourse Grammar. A typologically-based Theory of Language Structure*. Oxford: Oxford University Press.
- Jansen, Wim. 2013. Esperanto parts of speech in functional discourse grammar. *Linguistics* 51(3): 611–652.
- Jansen, Wim. 2015. Was ist und wozu brauchen wir eine Referenzgrammatik des Esperanto? [What is a reference grammar of Esperanto, and why do we need it?]. In Cyril Brosch & Sabine Fiedler (eds.), *Die Plansprachenbewegung rund um den Ersten Weltkrieg* [*The Planned Language Movement Around the First World War*]. Berlin: GIL, 97–107.
- Jansen, Wim. 2016a. Derivation in Esperanto. *Linguistics* 54(5): 1101–1133.
- Jansen, Wim. 2016b. *Ticklish Aspects of Esperanto Grammar*. Unpublished course guide for the John Buchanan Masterclass on Esperanto language, literature, and culture, University of Liverpool, 16–18 March, 2016. Updated and re-issued on 24 March 2016.
- Jansen, Wim. 2020. Pensoj pri la sufikso *-ec* en la Esperanta leksiko [Thoughts about the suffix *-ec* in the Esperanto lexicon]. *Language Communication Information* 14: 53–72.

- Kalocsay, Kálmán & Gaston Waringhien. 1935. *Plena Gramatiko de Esperanto* [Complete Grammar of Esperanto]. Budapest: Literatura Mondo. First edition.
- Kalocsay, Kálmán & Gaston Waringhien. 1985. *Plena Analiza Gramatiko de Esperanto* [Complete Analytical Grammar of Esperanto]. Rotterdam: UEA. Second edition.
- Koutny, Ilona; Cyril Brosch, Maciaj Jaskot & Horst Vogt. 2008. *Germana-Esperanta-Pola Etvortaro pri Lingvo kaj Komunikado* [German-Esperanto-Polish Pocket Dictionary of Language and Communication]. Poznań: proDRUK.
- van Oostendorp, Marc. 1999. Syllable structure in Esperanto as an instantiation of universal phonology. *Esperantologio. Esperanto Studies* 1: 52–80.
- de Saussure, René. 1969 [1911]. *Konkurso: kia estas la logika senco de l' vorto "grando"?* [Contest: What Is the Logical Meaning of the Word "grando"?]. Saarbrücken: Iltis. Photographic reprint.
- de Saussure, René. 1982 [1914]. *La vort-teorio en Esperanto* [Word Theory in Esperanto]. Saarbrücken: Iltis. Photographic reprint.
- de Saussure, René. 1969 [1915]. *Fundamentaj reguloj de la vort-teorio en Esperanto* [Fundamental Rules of the Word Theory in Esperanto]. Saarbrücken: Iltis. Photographic reprint.
- de Saussure, René. 1916. La vort-strukturo en Esperanto [Word structure in Esperanto]. *Esperanto*: 3–20.
- Waringhien, Gaston (ed.). 2002. *La Nova Plena Ilustrita Vortaro de Esperanto* [The New Complete and Illustrated Esperanto Dictionary]. Paris: SAT.
- Wennergren, Bertilo. 2005. *Plena Manlibro de Esperanta Gramatiko* [Complete Handbook of the Esperanto Grammar]. El Cerrito, CA: ELNA. Internet version: <https://bertilow.com/pmeg/index.html>, version 15.1 of 15 September 2020.
- Zamenhof, Lejzer. 2004 [1887]. *Международный языкъ. Предисловіе и полный учебникъ* [The International Language. Preface and Complete Textbook]. Moscow: Impeto. Photographic reprint of the Russian edition of Zamenhof's first brochure, published under the pseudonym Dr. Esperanto.
- Zamenhof, Lejzer. 1888. *Dua libro de l' lingvo internacia* [Second Book of the International Language]. Warsaw: Kelter.
- Zamenhof, Lejzer. 1963 [1905]. *Fundamento de Esperanto* [Foundations of Esperanto]. Marmande: E.F.E.

Wim Jansen

wimjansen@casema.nl