Esperantologio / Esperanto Studies **1** (1999), 52–80

Syllable structure in Esperanto as an instantiation of universal phonology

Marc van Oostendorp

The linguistic discipline of phonology is underrepresented within the field of Esperanto studies.[1] Most grammatical work concentrates on syntax and morphology, but the sound structure is ignored in many works on the grammatical structure of the language. A serious monograph discussing the most relevant aspects is yet to be written.[2]

One of the reasons for the relative lack of interest may be the fact that at first sight Esperanto does not have the type of phonological system that would excite phonologists. One of the official sixteen rules of Esperanto phonology (rule number 9, Kalocsay & Waringhien 1985:19) is:

> Every word is read aloud as it is written.

This statement is of course rather informal. If we translate it into the terminology of modern phonology, we could say that phonological elements do not alternate or get deleted: the orthographic representation gives us underlying structure and surface structure at the same time. Underlying vowels and consonants stay the way they are in every phonological context. The rule (if seen as a rule guiding language planning, rather than a descriptive device) therefore had as an effect that Esperanto does not have any interesting phonological alternations. Its morphology is completely agglutinative: there is no allomorphy, no fusion, and there are no assimilation or dissimilation rules. We sometimes find some discussion in the literature (Waringhien 1962, Wells 1978/1989) whether or not allophonic variation is permissible; e.g., whether or not ⟨n⟩ can be pronounced as a velar nasal in a word like *banko* 'bank'. Most of the discussion concerning this issue is prescriptive, rather than descriptive or theoretical, in nature. Besides, issues such as this do not seem very interesting from a theoretical point of view: every serious theory of phonology can describe the kind of place assimilation presumably attested here.

In this paper, I will try to show that the lack of interest displayed by esperantologists for phonology is as unjustified as the disdain for Esperanto displayed by phonologists. In my view, there is a lot to be learned about the sound structure of Esperanto by applying a modern methodology to it; and there is something to be learned about language in general by studying Esperanto.

---

[1] This article is partly based on an unpublished manuscript written in Esperanto (Van Oostendorp 1994). Thanks to Michael Redford for comments on a previous version.

[2] Cf. the parts on phonology in Kalocsay & Waringhien (1985) and Wells (1978/1989), and the studies by Kawasaki (1936–1953, 1961), Mangold (1985), Okamoto (1925) for questions that are relevant for the topics studied here.

Two remarks are in order here. First, in this article I confine myself to a branch of phonotactics that is more or less abstract, i.e., independent of the details of articulatory or acoustic phonetics. I study the phonotactic structure of words such as they appear in a dictionary, and show that generalisations can be made in this domain that are anything but trivial. This does not imply, of course, that phonetic investigations into the sound structure of Esperanto are impossible or undesirable. Yet it can be expected that the actual phonetic implementation varies widely among speakers with different mother tongues and language attitudes. I suppose that phonetic research into this variation could profit from a description of the aspects that all of these pronunciation have in common – the abstract phonological system.

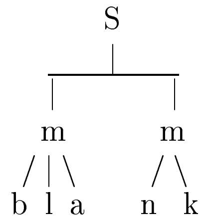
The second remark is directly related to the first. In order to describe classes of sounds, phonologists usually make use of a terminology that is derived from phonetics. I follow this practice here and write, e.g., about “alveolar”, “palatal” and “labial” consonants or about “high” vowels; yet I do not wish to imply that the most common or the preferred place of articulation for “alveolar” [t] is indeed at the alveolar ridge. Other places are attested and I do not wish to bestow myself with the authority to proclaim any particular pronunciation as either good or bad. Important for me is that [t, d, s, ...] form a natural class in the abstract phonology. I could have called this class *A*, but “alveolar” seems a better alternative from a mnemotechnic point of view. For lack of a good description of Esperanto phonetics, the reader is kindly requested to be cautious about other names of natural classes as well.

This article is structured in the following way. In section 1, I give a short introduction into syllable structure as applied to underived words in Esperanto. The syllable is divided into two subconstituents, the onset (discussed in section 2) and the rhyme (discussed in section 3). Section 4 deals with the complications of syllable structure in morphologically complex words. The last section is devoted to a conclusion.

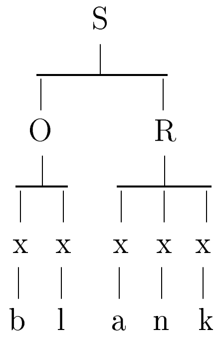
1. The syllable in underived words

The internal structure of the syllable has received quite a lot of attention in the phonological literature of the past few decades. Most theorists nowadays would agree that a syllable is not just an unstructured bunch of segments. The general agreement ends, however, as soon as one tries to be explicit about what the internal structure of a syllable is. There are two major schools of thought (cf. Blevins 1995). The differences between these schools are best illustrated by giving the respective structures that these schools would assign to the English syllable *blank*:

(1a)



(1b)



In mora theory, represented in (1a), a syllable (S) can consist of one or two moras (m): open syllables with a short vowel have one mora, closed syllables or syllables with a long vowel have two moras. In the latter cases, the second mora dominates the second half of the long vowel and the consonants following the vowel. Mora theory has been succesfully applied in the analysis of the interaction between syllable structure and stress (Hyman 1985, Hayes 1995, Kager 1995) and in the analysis of Prosodic Morphology phenomena such as infixation and reduplication (McCarthy and Prince 1986, 1993, 1995).

In onset-rhyme theory, represented in (1b), a syllable consists of an onset (O) and a rhyme (R).

The rhyme dominates the vowel and all consonants following it, while the onset dominates the consonants preceding the vowel. Onset-rhyme theory seems more succesful in describing the phonotactics of a language. Since phonotactics are the topic of the present contribution, and since Esperanto has a fairly simple stress rule which does not refer to syllable structure, and does not display any Prosodic Morphology phenomena at all,[3] I will use the onset-rhyme model in this article.

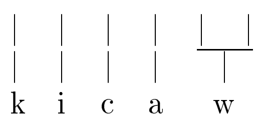
1.1. Long segments

The reader will note that the letters of the word *blank* (which of course represent phonemes or **segments** as I will call them) in (1b) are not directly attached to the onset and rhyme nodes. There is an intervening layer of **x-slots** which represent timing units (McCarthy 1979, Levin 1985). These x-slots are used to represent length of consonants and vowels: long segments are supposed to consist of one segment, attached to two x-slots.

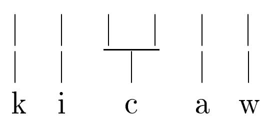
According to Kalocsay & Waringhien (1985), long vowels also play a role in Esperanto. We would basically find a vowel of this type in the penultimate syllable of a word (i.e., the syllable that carries primary stress) if this syllable is open. The contrast is therefore nonphonemic: it does not serve to distinguish between words and there are no minimal pairs of words with a different meaning and differing in form only in the length of one of the vowels. It seems therefore fair to say that vowel length is marginal in Esperanto at best.

Next to long vowels, languages can also use long consonants. Also these are represented as one (consonantal) segment attached to two different timing slots. An example of a language displayning genuine contrasts in consonantal length is Sierra Miwok, a Penutian language from California. In this language we see a contrast between the word forms *kicaww* and *kiccaw*. Both of them are forms of a verb meaning 'to bleed', but the first form approximately means 'will bleed', and the second 'bleeds continuously'. The syllable structure of these two forms is as follows:

(2a)



(2b)



[3]A possible exception are the hypochoristics formed with *-ĉjo* and *-njo* for male and female names respectively: *Vilĉjo* 'Bill' (< *Vilhelmo*), *panjo* 'mum' (< *patrino* 'mother'), etc. See section 2 for discussion.

The status of long consonants in Esperanto is a matter of debate[4]. There are a few minimal pairs involving geminate consonants morpheme-internally,[5] the most famous of which undoubtably is *finno* 'Finn' – *fino* 'end'. The number of these, however, is very small. Furthermore, morphological compounding can create sequences of identical consonants: *kapparto* 'part of head' (< *kapo* 'head' + *parto* 'part'). Words of the finno-type are discussed in Van Oostendorp (1998a, 1999); those of the kapparto-type in section 4. In both cases we will see that there is something special about these forms. This allows us to formulate the following principle about Esperanto sound structure:

Principle 1. Disallowance of long segments. *No segment may be linked to more than one timing slot.*

1.2. Complex segments: a first approach

Every segment corresponds to exactly one timing slot; one may therefore wonder whether the line with x-slots in the representation in (1b) is not superfluous in the description of Esperanto. There could, however, be another reason for still wanting to use this intermediate level of structure; this concerns the representation of the complex segments c, ĉ, ĝ, ĵ, ŝ and dz.[6] Learners of Esperanto have been wondering almost since the moment when the language came into existence whether or not these segments are one or two. Responding to a question on this issue, the creator of the language, Zamenhof (1927:34), has stated that:

> Your opinion on c, ĉ, ĝ is mistaken; it is true that *some* nations pronounce them as ts, tŝ, dĵ, but not *all* nations do this.

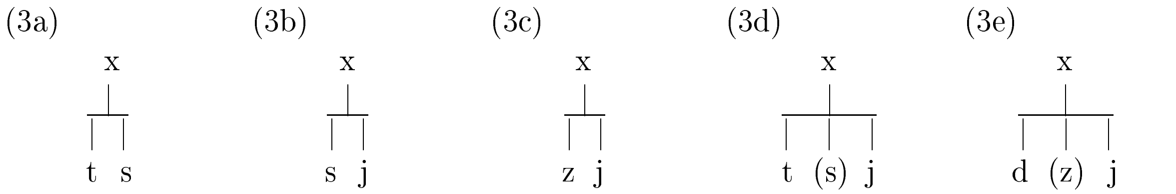
A similar line of reasoning could be assumed to apply to ŝ and ĵ, the voiceless and voiced fricative that could be seen by some as renderings of the sequences [sj] and [zj]. Kalocsay & Waringhien (1985) claim that this should probably be extended to ⟨dz⟩ and there are several reasons to adopt this position. The most important one is that otherwise ⟨c⟩ would be the only voiceless consonant that does not have a voiced counterpart in Esperanto.

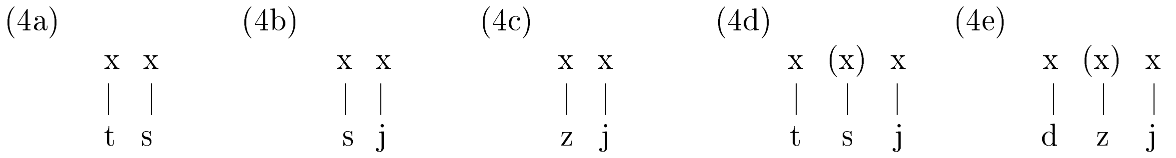
As we will see below (subsection 2.6), there is also a rather strong argument *not* to consider the dz as one segment, on a par with c, ĉ and ĝ: the latter segments can all freely occur at the beginning of a word, but this is not true for dz. The situation concerning this segment is therefore somewhat paradoxical. Using a separate tier of x-slots the similarities *and* the dissimilarities between long ⟨ts⟩ and short ⟨c⟩, *ŝ* and *sj*, etc., could be represented in the following way:

[4] Cf. Albault (1998) for an overview of all the relevant facts and a careful description of the possible phonetic realisations of biconsonantal graphemes.

[5] Morpheme-initially, there are no long consonants at all, to the best of my knowledge; **ssato* is not a feasible word of Esperanto.

[6] The symbol c stands for a coronal affricate in Esperanto orthography; ĉ for a voiceless palatal stop or affricate and ĝ for a voiced palatal stop or affricate; ĵ is a voiced palatal fricative and ŝ a voiceless one. All other Esperanto letters may be assumed to correspond to their IPA correlate for the purposes of this article.





Without further refinements, this type of representation still leaves us with a few problems. One of them is why we can combine [t] and [s] or [t] and [j] under one segment but not, e.g., [k] and [s] or [p] and [j]. Some further discussion of this will be provided in subsection 2.6.

In the representations of [ĉ] and [ĝ] above, I put the [s] and [z] in brackets. There is no reason for these to be present in the abstract phonological representation. We have no contrast between a monosegmental [tj] unit on the one hand and another monosegmental [tsj] on the other. Also, the phonological analysis gets slightly simplified if we assume [ĉ] to be monosegmental [tj] and [ĝ] monosegmental [dj], because it allows us to establish the following principle of Esperanto monosegments:

Principle 2. Principle of complex segments. *Esperanto allows the attachment of two segments to one timing slot if the first segment is an alveolar obstruent* ([t, d, s, z]) *and the second one* [s] *or* [j].

A special status is alotted here to alveolars; we will see below that this is a move we will have to make more often in our analysis.

According to this principle the following consonant combinations can be units of Esperanto: /ts/ (= [c]), /tj/ (= [ĉ]), /dj/ (= [ĝ]), /sj/ (= [ŝ]) and /zj/ (= [ĵ]). Apart from this it allows [ds], [zs] and [ss]. The first of these we could equate with ⟨dz⟩; the other two may be seen as indistinguishable from 'single' [z] and [s] respectively (a segment starting as a alveolar fricative and ending as a alveolar fricative simply *is* a alveolar fricative).[7]

2. The onset

I now turn to one of the two immediate subconstituents of the syllable. In the next section I discuss the rhyme, but in this section we start, of course, with the onset.

The simplest syllable type – the one we find in all languages of the world, the one that infants learn first – consists of exactly one consonant followed by exactly one vowel: Esperanto words as *po* 'at the rate of, each', *ne* 'no', *nura* 'mere', *rimo* 'rhyme' and *satelito* 'satellite' all fit into this template.

It is sometimes claimed in the literature that there are also languages which allow only this type of syllable and no others. It turns out to be rather difficult, however,

[7]If /ds/ = [dz] and /zs/=/zz/=[z], we should add to the Principle of complex segments a statement that the second segment harmonizes in voice with the preceding segment: an /s/ preceded by a voiced [d] or [z] turns into a voiced [z].

to find unambiguous examples of such languages. The West-African language Senufo may be a case in point (Clements and Keyser 1983) and languages spoken on the Fiji islands are also sometimes mentioned.

The group of languages in which the onset may also be empty seems much bigger, in particular if we only look at the first position of the word. Such languages allow for the type of sound combinations exemplified by the Esperanto words *amo* 'love' and *ataki* 'to attack'.

In some languages, empty onsets can only occur in the initial position of a word. At least a tendency to this effect can be observed in Esperanto as well. Many words start with a vowel, but there is only a very small group of (nonderived) words in which we can find such a syllable word-internally (*kaoso* 'chaos' is an example of such an exception).

2.1. A table of biconsonantal onsets

Apart from zero or one consonant, the Esperanto syllable can contain also two consonants, as in *prun.ti* 'borrow', *kom.pre.ni* 'understand' and *kna.bo* 'boy' (the dot indicates a syllable boundary). If the first segment is [s] or [ŝ], the onset can even contain three segments: *stra.to* 'street', *skla.vo* 'slave', *ŝtrum.po* 'sock', etc.; I will return to this below. For now, it is important to observe that not every combination of two segments can serve as an Esperanto onset. The first segment always has to be an element of the set {b, d, f, g, k, p, s, ŝ, t, v} and the second one an element of {r, l, n}. I provide a list of all possible combinations below:[8]

[br]	*bruna* 'brown', *brako* 'arm', *branĉo* 'branch'
[bl]	*blua* 'blue', *blago* 'joke', *bloko* 'block'
[bn]	*not attested*[9]
[dr]	*drinki* 'to drink (heavily)', *droni* 'to drown'
[dl]	*not attested*
[dn]	only in geographic names (*Dnepro, Dnestro*).[10]
[fr]	*franca* 'French', *fraŭlo* 'bachelor'
[fl]	*flava* 'yellow', *Flandrio* 'Flanders'
[fn]	*not attested*[11]
[gr]	*granda* 'big', *griza* 'grey'
[gl]	*glaso* '(a) glass', *gliti* 'to glide'
[gn]	*gnomo* 'gnome', *gnuo* 'gnu'[12]
[kr]	*kreteno* 'cretin', *krepo* 'pancake'
[kl]	*klera* 'learned', *klara* 'clear'
[kn]	*knabo* 'boy', *knedi* 'to knead'[13]

[8] My main sources have been Waringhien (ed.) (1987), henceforth PIV, and Kawasaki (1936–1953).

[9] Other consonant clusters with [b-] in PIV: [bj] in the geographic name *Bjalistoko/Bjelostoko*.

[10] Other consonant clusters with [d-] in PIV: [dv] in the geographic name *Dvino*; [dz] in the name of the Greek letter *dzeta* (also spelled *zeta*).

[11] Other consonant groups with [f-] in PIV: [fj] in *fjordo* 'fjord'; [ft] in *ftizo* 'phthisis' a.o.

[12] Other consonant groups with [g-] in PIV: [gh] in *ghetto* 'ghetto'; [gv] in *gvidi* 'to guide', *gvati* 'to keep watch' a.o.

[13] Other consonant groups with [k-] in PIV: [kj] in the name *Kju*, [ks] in *ksantelasmo* 'xantelasm', [kv] in *kvar* 'four', *kvereli* 'to argue' a.o.

[lr]	*not attested*
[ll]	*not attested*
[ln]	*not attested*[14]
[mr]	*not attested*
[ml]	*not attested*
[mn]	*not attested*[15]
[nr]	*not attested*
[nl]	*not attested*
[nn]	*not attested*[16]
[pr]	*preni* 'to take', *profiti* 'to profit'
[pl]	*plano* 'plan', *plori* 'to weep'
[pn]	*pneŭmonio* 'pneumonia'[17]
[sr]	*not attested*
[sl]	*slipo* 'index card', *slango* 'slang'
[sn]	*snobo* 'snob', *snufi* 'sniff'[18]
[ŝr]	*ŝraŭbo* 'screw', *ŝranko* 'cupboard'
[ŝl]	*ŝlifi* 'to grind, to polish', *ŝlosilo* 'key'
[ŝn]	*ŝnuro* 'line'[19]
[tr]	*trajno* 'train', *tri* 'three'
[tl]	*not attested* (except in *tlaspo* 'thlasp')
[tn]	*not attested*[20]
[vr]	*vrako* 'wreck', *vringi* 'to wring'
[vl]	in the names *Vladimiro* and *Vladivostoko*
[vn]	*not attested*[21]
[zr]	*not attested*
[zl]	*zloto* 'zloty'
[zn]	*not attested*

The exceptions mentioned in the footnotes to this table are discussed in Van Oostendorp (1998a, 1999). Here, I will concentrate on the "regular" patterns.

2.2. Sonority

One way of describing the relation between the first and the second segment of the onset invokes a principle of sonority (cf. Clements 1990, Blevins 1995 and references cited there):

[14] Consonant groups with [l-] in PIV: [lh] in the geographic name *Lhaso*, [lj] in *ljamo* 'llama' (also spelled *lamo*) and the geographic name *Ljuŝun*, [lv] in the geographic name *Lvovo*.

[15] Consonant groups with [m-] in PIV: [mj] in *mjelo* 'spinal marrow'.

[16] Consonant groups with [n-] in PIV: [nj] in the geographic names *Njasa, Njemeno*.

[17] Other consonant groups with [p-] in PIV: [pf] in *pfenigo* '(German) pfennig', [ps] in *psalmo* 'psalm', *pseŭdo-* 'pseudo', [pŝ] in *pŝento* 'pschent', [pt] in *pterido* 'bracken', *ptialazo* 'ptialine'.

[18] [s-] can also be combined with [t-], [p-], [k-], etc. This has not been explicitly marked in this table.

[19] [ŝ-] can also be combined with [t-], [p-], [k-], etc. This has not been explicitly marked in this table.

[20] Other consonant groups with [t-] in PIV: [tb] in the geographic name *Tbiliso*, [tj] in the interjection *tju* and the geographic name *Tjurko*, [ts] as an interjection and in *tsetseo* 'tsetse' (according to PIV, this should actually be *ceceo*).

[21] Other consonant groups with [v-] in PIV: [vj] in *vjelo* 'hurdy-gurdy', *vjolo* 'viola'.

Principle 3. Sonority Sequencing Principle, Onset *(SSP). Between any member of a syllable and the syllable peak, a sonority rise or plateau must occur.*

Sonority is a technical term, defined (for example) in the following way:[22]

(5) **Sonority.** A segment S_1 is more sonorous than a segment S_2 if S_1 is more to the left than S_2 on the following scale:

vowels	*glides*	*liquids*	*nasals*	*obstruents*
a, e, i, o, u	j, ŭ, v(?)	l, r	m, n	p, b, t, d, k, g, f, v(?), s, z, c, dz, ĉ, ĝ, ŝ, ĵ

Vowels can only occur in the syllable rhyme, not in an onset, in Esperanto as well as in most (maybe even all) other languages. Also the appearance of glides in the onset is severly restricted. The [j] can only appear in that position if it is not preceded or followed by any other consonant (as in *jes* 'yes' and *jaro* 'year'). There is only a small set of words in which it can be the second segment (*fjordo* 'fjord'); I have argued elsewhere (Van Oostendorp 1998a, 1999) that the words in this set are "loanwords" from a phonological point of view: they have a more complex structure than the "core" Esperanto lexicon.

The distribution of the back rounded glide [ŭ] is even more restricted. It is almost a normal vowel in its preference for the rhyme. The only few exceptional cases in which [ŭ] occurs is in the onset; these again arguably are "loanwords" in a technical sense: *ŭato* 'watt'.

2.3. The status of [v]

It is also important to note some theoretical problems regarding the sonority of [v]. Assuming this segment is neither a nasal (because there is no nasal airflow), nor a vowel or a liquid, it could be described as either a glide or an obstruent.

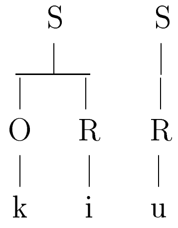
It is possible to provide arguments for both positions. On the one hand, [v] can occur in an onset immediately following [k] or [g], as in *gvidi* 'to guide' and *kvar* 'four'. [k] and [g] are obstruents and therefore the principle of sonority just outlined forces us to assume that [v] is more sonorous than an obstruent. That is not a problem in principle, since there are many languages in which we can find glides that are very similar to the Esperanto [v].

On the other hand, glides are often assumed to be closely related to a vowel. Such a relation clearly exists between [j] and [i], which are phonetically very similar: as a matter of fact, they are virtually identical, except that [i] appears in vocalic positions (the nucleus) and [j] in consonantal positions (the onset, and maybe the coda). This strong relationship can also be observed in words such as *kiu* 'who' and *piano* 'piano'. Those words are often pronounced approximately as [kiju], [pijano].[23] This fact can be understood if we assume that [j] in fact *is* an /i/ , occuring in consonantal position. Let us consider the word *kiu* for a moment (I will leave out the skeleton in order to not unnecessarily complicate the discussion):

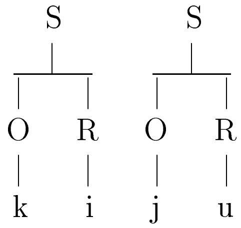
[22] This of course is an abstract and arbitrary definition, which may be replaced by something more phonetic (level of obstruction in the mouth) if one wishes.

[23] This observation has been made already by Zamenhof (1927).

(6)



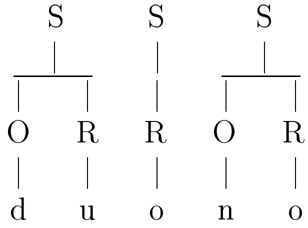
(7)



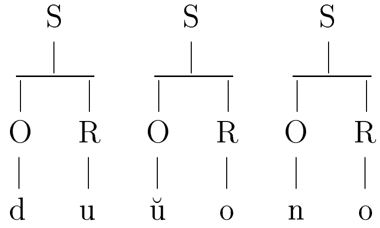
In (6) the second syllable does not have an onset. This is an undesirable situation. The system therefore tries to find a segment which can play that role. It then finds the /i/ which appears as [j] in consonantal position, as I just described. We then end up with the structure in (7).

The same line of reasoning explains the appearance of [ŭ] in /duono/–[duŭono]. We start out with the structure in (8):

(8)



(9)



In (8), the second syllable is looking for an onset and finds one in the form of /u/, which is represented by [ŭ] in consonantal position in (9) (apparently, even the fact that [ŭ] is rare in underlying onset position does not prevent its surfacing here). The only thing that matters is the material *preceding* the syllable. One never says *[kiŭu] or *[dujono].[24]

The problem with /v/ now is that it cannot correspond to any real vowel. It is phonetically most similar to [u], but the glide corresponding to that vowel is [ŭ], as we just saw. This is also indicated by the fact that we can find some (albeit not many) words where [v] precedes a liquid: *vringi* 'wring', *Vladimiro* 'Vladimir'. From this I tentatively conclude that /v/ is an obstruent after all; the issue that it nevertheless can be preceded by [g] and [k] will be taken up shortly below.

Apart from the glides [ŭ] and [j], there are other consonants that prefer to stay on their own in a syllable onset, notably [m], [h, ĥ] and almost all "complex" consonants ([c, ĉ, ĵ, ĝ] but not [ŝ]). Also these will be discussed in somewhat more detail below.

2.4. Phonological government

It is convenient to introduce now some terminology in order to describe these restrictions. The most important term here is government.[25] Phonological government could

[24] One of the anonymous reviewers suggests that this may not be true for everybody: Korean speakers of Esperanto frequently say, e.g., [unuje, duje] for *unue* 'firstly', *due* 'secondly'. This may therefore be a case of authentic phonological variation. After [e] one can frequently observe a [j]: /teo/-[tejo] 'tea', and after an [o] we may find [ŭ]: /poeto/-[poŭeto] 'poet'. There is no glide corresponding to /a/ and one also does not say *[kajoso] or *[kaŭoso] for /kaoso/ 'chaos'.

[25] The term derives from traditional grammar, where one says that some verbs or prepositions *govern* a certain case. For example, the directional prepositions govern the accusative in Esperanto, but the locatives govern the nominative. The term was taken up in the so-called *Government and Binding* framework of generative grammar in the 1980's (Chomsky 1981). Some phonologists, notably Kaye

be defined as follows:

(10) *Government:* In a subsyllabic constituent (onset or rhyme), every segment governs the element to its right.

This definition states that the /f/ in the first syllable of *franca* 'French' governs the /r/ and the /a/ in that syllable governs the /n/. We can now state the sonority sequencing principle in terms of government:

(11) *Sonority Sequencing Principle* (to be revised).

a. In the onset, the governing segment is less sonorous than the governee.

b. In the rhyme, the governing segment is more sonorous than the governee.

The observations on possible onsets just made can now be formulated as in (12):

(12) a. Complex segments and [h, ĥ] cannot govern.

b. The /j/ and /m/ can neither govern nor be governed in the onset.

c. The /ŭ/ cannot occur in an onset at all.[26]

With these extra restrictions, the Principle of Sonority gives us all the possible combinations of consonants in Esperanto, at the same time excluding, e.g., those in (13):

(13) a. *rbumo, *nguo, *lkera

b. *bbumo, *gguo, *kkera

In the initial onsets of the words in (13a), the first consonant is more sonorous than the second one. In the words in (13b), both consonant are equally sonorous. Both situations are disallowed by Sonority Sequencing.

Nevertheless there are still two groups of problems which the SSP does not address. First, if a mere difference in sonority level would suffice, we should be able to find words starting with sequences such as [nr-]. The fact that such forms do not exist can only be described if we sharpen the definition of the SSP: there should not just be a difference in sonority between governing segment and governee; this difference should be maximal. In order to do this, we may assign a number to every level of sonority: vowels get 5, glides 4, liquids 3, nasals 2 and obstruents 1. We can now let the SSP refer to these sonority levels:

Principle 4. Sonority Sequencing Principle, Onset (revised). *If we subtract the sonority level of the governor in the onset from the sonority level over the governee, the result should be larger than 1.*

et al. (1985), have subsequently introduced the term into phonological theory as well. The popularity of 'government' as a theoretical tool has diminished both in syntax and in phonology in the past few years; my primary motive for using it here is descriptive convenience.

[26] This statement needs to be qualified a little bit; as we have just seen, ŭ can occur in the onset of a syllable if it occurs in the rhyme of the preceding syllable at the same time ([duŭone]). Secondly, there is a small set of words which have ŭ in onsets, e.g., *ŭato* 'Watt', *sŭahila* 'swahili'.

According to this definition, [tr], [kl], etc. are possible onsets, because the sonority of [t] is 1 and that of [r] and [l] is 3; we note that $3 - 1 = 2$ and 2 is larger than 1. This does not hold for [nr]. The sonority of [n] is 2 and that of [r] is 1; so the sonority difference between the two segments is only 1, which is too small.

We now have a new problem if we start considering [kn-] in *knabo* 'boy'. In that word, the calculation is as follows: the value of [k] is 3 and the value of [n] is 2, so the difference is only 1. That would mean that a common Esperanto word such as *knabo* is not well-formed. This result is clearly absurd.

If we reconsider the table of words in section 1, and if we put apart the words starting with [s-] or [ŝ-], which will be discussed in section 2.5, we see that only the velar plosives [k] and [g] can appear before an [n]. Those two velar obstruents evidently are much stronger than the other segments; they are exactly the segments that also can occur before [v] in *kvar* 'four' and *gvidi* 'to guide'. This fact can be made explicit in the following way:

(14) Velar plosives can exceptionally govern [n] and [v].

Why it should be exactly [n] and [v] that can enter into this kind of exceptional relationship, or why it should be exactly the velar plosives that can govern these elements, is unclear to me. I leave this open for future research.[27]

Even now our description of the bisegmental onset is not complete, however. We still have to explain why [tl-] and [dl-] are impossible groups of consonants, even though they conform to the most precise version of the SSP just given. Now the point about these clusters seems to be that [t], [d] and [l] are all three so-called "alveolar" consonants: they are pronounced by pressing the tip of the tongue against the alveolar ridge, just behind the teeth. We may conjecture, then, that although the segments in [tl-] and [dl-] are sufficiently different in terms of sonority, they are not sufficiently different in terms of place of articulation.[28] In other words, it looks as if the SSP is just an instantiation of a more general principle of syllable structure in Esperanto:

Principle 5. Principle of Maximal Differentiation. *The segments of subsyllabic constituents (onsets and rhymes) have to be maximally different.*

This principle is not unique to Esperanto, of course. Since most of the morpheme inventory of the language derives from Indo-European languages, so do most of the phonological restrictions on possible morphemes. Besides, many non-Indo-European languages seem to obey a similar restriction.

2.5. The special behaviour of [s] and [ŝ]

I should now say a few words about [s] and [ŝ]. A quick look at PIV reveals that almost all consonant groups that can be bisegmental onsets also can be preceded by [s] and,

[27] One of the reviewers points out that there is a small class of words starting with [pn-]: *pnemonio* 'pneumonia', *pnematika* 'pneumatic'.

[28] Note that this implies that [r] should not count as an alveolar consonant in Esperanto from a phonotactic point of view, even if it is phonetically pronounced at a position very close to the alveolar ridge by many speakers: there is no evidence that those speakers cannot pronounce [trajno]. Phonetically this seems certainly justified, since the rhotic liquid can be articulated in a variety of positions, many of them non-alveolar.

somewhat less regularly, by [ŝ]. The reader should compare these lists of words with those in the next table:

[sĉ, sd, sg, sĝ, sh, sĥ, sj, sĵ, sr, ss, sŝ, sz]	*not attested*
[sb]	only *sbiro*
[sc]	*scii* 'to know', *sceno* 'scene'
[sf]	*sfero* 'sphere', *sfinkso* 'sphynx'
[sk]	*skandi* 'scan', *skemo* 'scheme'
[sl]	*slango* 'slang'
[sm]	*smokingo* 'dinner-jacket'
[sn]	*snobo* 'snob', *snufi* 'sniff'
[sp]	*spezo* 'turnover', *sperta* 'expert'
[st]	*stelo* 'star', *stafeto* 'courier'
[sv]	*svingi* 'to swing', *svelta* 'svelte'
[sbr, sbl, sdr, sfr, sfl, sgr, sgl, sgn, skn, ssr, ssl, sŝl, sŝr]	*not attested*
[skr]	*skribi* 'to write', *skrupulo* 'scruple'
[skl]	*sklavo* 'slave'
[skv]	*skvamo* 'scale of fish'
	skviro 'esquire' (usually: *eskviro*)
[spl]	*splito* 'splinter', *splisi* 'to splice'
[str]	*strato* 'street', *striko* 'strike'
[ŝb, ŝc, ŝĉ, ŝd, ŝf, ŝg, ŝĝ, ŝh, ŝĥ, ŝj, ŝs, ŝŝ]	*not attested*
[ŝk]	*ŝkopi* 'to bail'
[ŝl]	*ŝlosilo* 'key', *ŝlimo* 'slime'
[ŝm]	*ŝmiri* 'to smear', *ŝminko* 'grease-paint'
[ŝn]	*ŝnuro* 'line'
[ŝp]	*ŝparo* 'savings', *ŝpuro* 'gauge (of track)'
[ŝr]	*ŝranko* 'cupboard', *ŝraŭbo* 'screw'
[ŝt]	*ŝtato* 'state', *ŝteli* 'to steal'
[ŝv]	*ŝviti* 'to perspirate', *ŝvebi* 'to float'
[ŝbr, ŝbl, ŝdr, ŝfr, ŝfl, ŝgr, ŝgl, ŝkl, ŝkr, ŝpl, ŝsl, ŝsr, ŝsn, ŝŝl, ŝŝr, ŝŝn]	*not attested*
[ŝpr]	*ŝpruci* 'to spray', *ŝproso* 'sprout'
[ŝtr]	only *ŝtrumpo* 'sock'

From this we can conclude at least that [s] and [ŝ] can only be followed by those clusters of consonants which themselves are possible onsets:

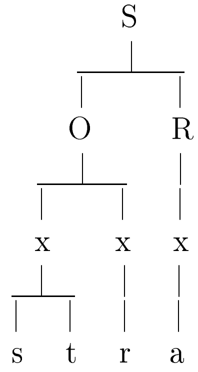
(15) If there is an Esperanto word that starts with [s+X] or [ŝ+X], X being a consonant cluster, then there also is an Esperanto word that starts with [X].

Another observation, which we can make immediately, is that clusters starting with [s-] or [ŝ-] violate almost every principle that we have established until now. For instance, the cluster [st-] violates the SSP because the sonority difference between [s] and [t] is 0; it violates the more general Principle of Maximal Differentiation also in another way, because [s] and [t] are both alveolar consonants.

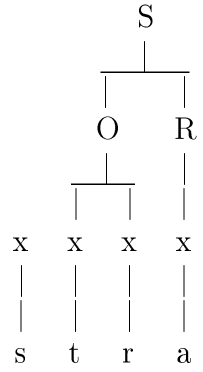
2.5.1. The syllable position of s and ŝ

According to some phonologists, the deviant behaviour of [s] and [ŝ] possibly indicates that these segments are not real parts of the onset. The Esperanto onset in that case would consist of maximally two positions, but (at the beginning of the word) it could be preceded by [s] or [ŝ]. This idea can be worked out in at least two different ways. Some linguists (e.g., Selkirk 1982 for English and Noske 1988 for French) propose that [st], [sk] etc. should be represented as complex segments. They would give the structure in (16a) to the first syllable in *strato* 'street'. Other linguists (e.g., Van Oostendorp 1995 for Dutch) propose that [s] and [ŝ] should stay outside the syllable structure of the word altogether. In that case the first syllable of *strato* would have the structure of (16b):

(16a)



(16b)



I think that also in Esperanto the majority of arguments is in favour of the second type of structure. For let us suppose for a moment that the first structure would be the right one. That would imply that:

1. we would need to redefine the Principle of Complex Segments in a way that is much less elegant;

2. the complex segments [st], [sk], [ŝk], etc. would be able to govern another complex segment, unlike the complex segments c, ĉ, etc. (compare for instance *strato* with *crato);

3. Esperanto would have complex segments with at least three elements on the melodic tier (e.g., in *ŝpari* 'to save'). This would make Esperanto a rather marked type of language because those kinds of 'very complex' segments are rather rare cross-linguistically.

Compared to this complicated state of affairs, the structure in (16b) seems much simpler.

The problems mentioned above can be solved in the following way:

First, we do not need to change the Principle of Complex Segments, because no complex segments are involved in this structure. On the other hand, of course we would need to change a principle that has not been made explicit until now.[29] This is Junko Itô's (1986) Principle of Prosodic Licensing:

[29] Of course, since [s] and [ŝ] are outside the syllable, they are also not subject to Principle 4.

Principle 6. Principle of Prosodic Licensing. *All elements in a phonological structure need to be licensed by incorporation in some larger phonological structure (except for licensed elements at the periphery).*

The Principle of Prosodic Licensing states that every element of the skeleton needs to be in an onset or a rhyme; every onset and rhyme needs to be in a syllable; every syllable needs to be in a stress foot; every foot needs to be in a phonological word; and every word needs to be in a phonological phrase.

The [s] and [ŝ] clearly are exceptions to this principle, if we assume that they stay outside of the syllable structure proper. According to the Principle of Prosodic Licensing, this is allowed, because they are usually in a peripheral position in the word.

Apart from this, they may get licensed by a special mechanism. In order to see this, we need to study the internal structure of these two consonants. The distinctive features of vowels and consonants can be divided according to several criteria, but one of them is the place of articulation: [p, b, f, m] are *labials*, because they are pronounced at the lips, [k, g] are *velars*, because they are pronounced at the velum.

Apart from the labial and the velar place of articulation, we also have the alveolar place of articulation, where the alveolar consonants [s, t, l, n] are pronounced. Now it is widely agreed upon that there is something special about the alveolar place of articulation: this arguably is the simplest place to articulate a consonant, because it is easier to move the tip of the tongue than to move the back of the tongue or the lips.[30]

This implies that we do not need to specifically indicate the place of articulation in the mental representation of a sound, if this sound is an alveolar. We need to remember that [ĥ] is a velar and [f] a labial, but [s] does not need to be specified for a place: it is pronounced at the simplest place available to humans, i.e., the tip of the tongue (and similarly for [ŝ], although this segment is of course slightly more complex).

Now some phonologists (e.g., Itô & Mester 1993) have proposed that the Principle of Prosodic Licensing should not target segments, but rather places of articulation. We cannot discuss Itô's and Mester's arguments here. It should be clear, however, that if these linguists are right, the alveolars do not need to be incorporated in the syllable structure. On the other hand, it is still not clear why it should be exactly the voiceless fricatives [s] and [ŝ] that occur in this position, and not the other alveolar consonants of Esperanto [z, ĵ, d, t]. The same seems to be true in many languages of the world.

Another argument for assuming that [s] and [ŝ] are outside the syllable is that the government relation in *strato* in that case no longer needs to concern us. [tr] is a well-formed Esperanto onset; this is not affected by whether or not [s] precedes it.

On the other hand, more can be said about this as well. Although everything which follows [s] or [ŝ] is a well-formed onset, not every well-formed onset can follow [s] or [ŝ]. A particularly strong generalisation, for instance, is that clusters with *voiced* obstruents (such as [b, d, . . .]) are disallowed:

(17) After [s] and [ŝ], voiced obstruents are disallowed.

This probably should be related to the fact that [s] and [ŝ] themselves are voiceless. Clusters of obstruents in which one is voiceless and the other voiced are very rare in

[30]See Paradis and Prunet (1991) for more phonetic and phonological arguments why alveolar is "special" in the sound system of human languages.

Esperanto, as well as in other languages. For instance, we have *aktoro* 'actor', but not **akdoro* or **agtoro*. (The few counterexamples *ekzemplo* 'example', *ekzisti* 'exist' for one reason or another involve [kz].[31]

Another observation we can make here is that after [s] and [ŝ] we do not find the clusters [kn], [gn] and [gv], and we find only one common word with [kv] (*skvamo* 'scale of fish'). We have seen that these onset clusters are also special if they start the word (the sonority difference between the two consonants involved is too small). This means that we only find 'normal' onset clusters after [s] and [ŝ] .

As I have already pointed out several times above, it is not unusual that exceptional structures prefer to occur at the edges of words. This seems to be the case also here: velar obstruents can only license other consonants if they are in an absolute word-initial position.

Another gap in the table is more mysterious to me: the fact that after [ŝ] we do not find clusters with [l]: why are there no words starting with [ŝkl], [ŝpl], etc.? I do not have a good explanation for this gap; maybe it is only accidental.[32]

It would be worth trying to find out whether speakers of Esperanto would accept, e.g., *ŝplito, ŝklado* as possible words of Esperanto.

All in all, it seems to me, then, that we should preferably view word-initial [s] and [ŝ] as extrasyllabic segments.[33]

2.6. More on complex segments

To finish this discussion of the onset of the Esperanto syllable, I want to briefly return to the structure of the complex segment in Esperanto. As pointed out above, the first Esperanto scholar, L. L. Zamenhof, claimed that [c] really was different from [ts] and [ŝ] from [sj]. Now that we know more about syllable structure, there might be a few reasons to think that these differences are not all that large.

In the first place, it should be observed that [tj] and [ts] themselves are not possible consonant clusters: they occur neither morpheme-internally, nor in the initial (or final) position of a morpheme: there are no words such as **petjo* or **patso* or **tjalko* and the only word with initial [ts] in PIV, *tsetseo*, is called a "misspelling" for **ceceo* 'tsetse'. If we suppose that [c] is different from [ts], [ĉ] from [tj], etc., this is a problem. If we suppose that [c] is the usual way of writing [ts], and [ĉ] the usual way of writing [tj], on the other hand, this fact can be readily understood.

Even more important in this respect is the observation that complex segments never govern other segments in the rhyme, even though phonetically they are clearly

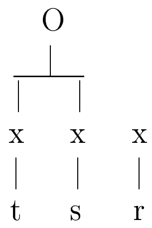
[31] It would be interesting to be able to experimentally verify whether such words are indeed pronounced as proscribed, or whether people allow themselves to say, e.g., [eksemplo] or [egzemplo] in fluent speech.

[32] Michael Redford (personal communicaton) points out that the explanation for this gap may be phonetic; [ŝ] probably has a partly retroflex pronunciation and in any case it will cause turbulence more or less in the central part of the oral cavities, whereas pronouncing [l] causes turbulence at the lefthand and righthand peripheries of this cavity. This may be too complex for a speaker to pronounce or for a listener to perceive correctly. Notice, however, that there are a few quite common words where [ŝl] occur next to one another (*ŝlosilo* 'key').

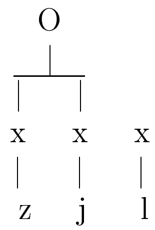
[33] Notice that we now incorrectly predict clusters such as *[sŝr] to be possible, since [ŝr] seems to be a well-formed onset. There seems to be a restriction on [s] and [ŝ] occuring next to one another, however. This may be kindred to the special relation both [s] and [ŝ] seem to entertain to complex segments.

obstruents: there are no words such as *[craro] or *[ĵleka]. (Of course we have to disregard the extrasyllabic [ŝ] here.) This can be understood if we suppose that those segments *themselves* occupy the two positions in the syllable. In that case, there is no longer an extra position for the liquid:

(18a)

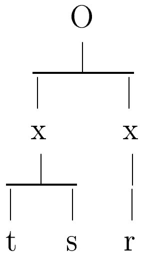


(18b)

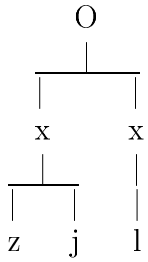


On the other hand, if complex segments are just one segment, there is no apparent reason why they would not tolerate another consonant after them. No principle discussed so far could explain why we do not have:

(19a)



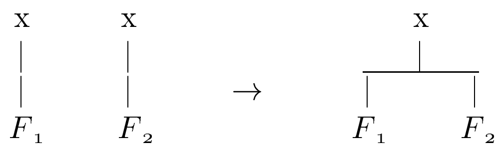
(19b)



On the other hand, the governing relations in (18) themselves are somewhat unusual: in the first tree, we have an obstruent governing another obstruent (violating the principle of sonority) and within the second tree the [j] is governed.[34] For this reason, I accept the analysis of complex segments as single segments. The problems just mentioned should then be solved in one way or the other. One way of doing this would involve the following two assumptions.

In the first place, we need to assume a rule that simplifies consonant clusters wherever this is possible:

(20)



Wherever this rule found the sequence of segments [ts], it would change this to the simple segment [c]. On the other hand, the sequence [ps] cannot be 'simplified' in the same way, since the corresponding single segment would be prohibited by the Principle of Complex Segments.

Secondly, until now we assumed that every onset has two *skeletal* positions. This assumption could be slightly changed; we might suppose that every onset has at most two positions (cf. Kager 1989 about a similar notion of *melodic complexity* used within

[34]Notice, however, that in both cases, this kind of governing relation is not impossible and occurs with other obstruents as well: *psikologo* 'psychologist', *fjordo* 'fjord'. In these other cases, however, we typically deal with words that could be classified as "learned"; see Van Oostendorp (1998a, 1999).

the rhyme). Because onsets can have at most two different *melodic* elements, this explains why we do not have *[cr], *[ĵl], etc.

Yet no matter whether we choose the monosegmental or the bisegmental representation for complex segments, the problem remains that complex segments also are not allowed after [s] and [ŝ] (except for [sc] in *scii* 'to know' and a few related words): there are no words such as *[sĉii] or *[ŝcii]. Intuitively, these clusters seem too complex; but within the system used here, there is no way to account for this complexity: if [s] and [ŝ] otherwise count as independent from the onset, why should they matter here?

Another problem is that I have found it convenient to follow Kalocsay & Waringhien (1985) in my analysis above in claiming that [dz] is the voiced counterpart to [c]. Unfortunately, however, apart from the name of the Greek letter *dzeta*, there are no Esperanto words starting with [dz-], even though all other complex segments can be found in that position quite frequently. Because complex segments cannot occur at the end of the word, as we shall see below, this means that [dz] can only occur in an intersyllabic position, where it could be analysed as occupying two positions (e.g., *edzo* 'husband' might be analysed as [ed.zo]). The reasons for assuming that [dz] is a complex segment on a par with [c] are rather weak, after all. On the other hand, if [dz] is not a complex segment, there is no voiced counterpart to [c] and this would be the only gap in the otherwise perfectly symmetrical consonant system of Esperanto. I have no solution to this paradox.[35]

3. The rhyme

Let us now turn to the other half of the Esperanto syllable, the rhyme. We immediately discover a methodological problem that we avoided while studying the syllable onset. In the previous section, I silently assumed that we can learn all there is to know about the syllable if we only look at the first syllable of the word: if something is a possible syllable, then it should be the first syllable of some Esperanto word. The advantage of this is that we can simply use the alphabetic order of the dictionary in order to find all the forms that needed to be considered. As long as we do not have an electronic version of PIV, the most extensive dictionary of Esperanto, this methodology cannot be used for studying the rhyme.

Apart from this, it does not make sense to study the rhymes of the last syllables of words when studying Esperanto phonology. The reason for this is that most Esperanto words end in a grammatical vowel; the number of these endings is small and there probably are more possible rhymes than possible grammatical endings.

We can now take one of two options:

1. We can postulate that we only consider morphemes, not words. Starting from the words *pilko* 'ball', *marŝi* 'march' and *kudri* 'to sow', and ignoring the grammatical endings *o* and *i*, we would have to accept *-ilk*, *-arŝ* and *-udr* as possible rhymes. In my view, this conclusion is hardly acceptable, since these stems in the normal case *are* followed by a grammatical morpheme, which furthermore always starts with a vowel. At least one of the consonants is syllabified with that vowel and therefore does not need

[35]One of the anonymous reviewers suggests that a phonetic test could be provided for those speakers that lengthen vowels in stressed open syllables. Those speakers would lengthen the [e] in *peco* 'piece', because [c] would be monosegmental, but not the [e] in *kverko* 'oak'. The testcase would be whether or not these speakers lengthen [e] in *edzo* 'husband'.

to be in the rhyme at all. – Yet there is a kind of tradition within Esperanto studies which seems to support this view: Zamenhof (1927) claimed that every morpheme, including grammatical endings, is to be considered an independent word.

2. Another possibility is to assume that the basis for syllabification is the morpheme plus a grammatical ending. I hope to show that this assumption gives us a much simpler view of the Esperanto rhyme. – Of course, this position is not necessarily in conflict with the one proposed by Zamenhof (and other Esperanto scholars) if we take a multidimensional view of linguistic structure. Things may be organized in a different way in the phonological plane than in the morphological plane, and Zamenhof was clearly talking about the latter.

In this section, I will use the following methodology. First, I will discuss word-final clusters. Secondly, I will discuss the word-internal clusters in "simple words", i.e., words that contain only one basic morpheme plus a grammatical ending. Section 4 will deal with the phonological structure of complex words.

3.1. Word-final clusters

Two classes of segment sequences are of interest to us now: the grammatical endings and the limited set of words (function words, or words in a closed class such as numerals) which do not have a grammatical ending:[36, 37]

Grammatical endings	*-a, -aj, -ajn, -am, -an, -as, -aŭ, -e, -el, -en, -es, -i, -in, -is, -o, -oj, -ojn, -on, -om, -os, -u, -us*
Closed class items	*unu, du, tri, kvar, kvin, ses, sep, ok, naŭ, dek, cent, mil; el, al, ĉe, da, de, dum, ekster, el, en, far, ĝis, inter, je, krom, kun, per, plus, po, por, post, preter, pri, pro, sen, sub, super, sur, tra, trans;* *ĉar, do, kaj, nek, sed, tamen;* *ke, kvankam, se; ajn, nur, eĉ, des, tuj, jes, ne, nu, ek, la*

The number of possibilities for a word to end in two consonants is very limited. I take this as an indication that the rhyme obeys to a restriction very similar to that imposed on the Esperanto onset:

(21) The rhyme contains maximally two segments.

There are two classes of apparent exceptions to this generalisation: the forms in which the vowel is followed by [-jn] (-ojn, -ajn: the accusative plural of nouns and adjectives, respectively) and the forms in which the second consonant is [s] or [t] (e.g., *post* 'after', *cent* 'hundred').

It is of course very significant that in the latter class of words the 'extra' consonant is always a voiceless alveolar [s] or [t]. We have already seen that voiceless alveolar fricatives can be extrasyllabic at the beginning of the word. At the end of the word, the

[36] I have put *-aŭ* in the set of grammatical endings in this table, even though it is not clear what grammatical category it would denote (it is used in some prepositions and some adverbs and in those cases it often cannot be elided in the same way as other grammatical endings). I do this for convenience only; no specific ideas about the grammatical function of this diphthong is at stake.

[37] I left out all interjections, because, as in many other languages, these do not seem to conform all rules of syllabification.

restriction apparently is slightly different but still very similar: here voiceless alveolar consonants that are not complex can be extrasyllabic.

Of course, the [n] is also alveolar; it is voiced, but redundantly so (Esperanto does not have voiceless nasals). Also this segment therefore fits into this category, and therefore also the *-ajn, -ojn* cases can be understood in this way.[38]

Another observation is that sonority is relevant in the rhyme as well. If we analyse *all* alveolars as extrasyllabic – so including those in *-as*, *ses* 'six', *jes* 'yes', etc. – there is only a handful of words with an obstruent in the rhyme. We have one word with [-p] (*sep* 'seven'), one with [-b] (*sub* 'under'), one with [-ĉ] (*eĉ* 'even') and four with [-k] (*ok* 'eight', *dek* 'ten', *nek* 'nor', *ek* 'let's start!'). Three out of these six exceptions are numerals; these have an exceptional syllable structure in many languages.[39] The other group of exceptions is sufficiently small to be able to claim that:[40]

(22) Only sonorants and vowels can appear in the rhyme.

We can be somewhat more precise than this: the first position of the rhyme is always occupied by a full vowel, the second position by a glide or a sonorant consonant. Also this sequence is the consequence of the SSP. In the onset, the least sonorous element was the leftmost; in the rhyme this is the most sonorous element:

(23) In the rhyme, vowels are the only possible governing elements; sonorant consonants and glides the only possible governees.

3.2. Similarities between the rhyme and the onset

Also in other respects, the rhyme is the mirror image of the onset. We have already seen that vowels will never appear in the onset and glides do so only exceptionally. Within the rhyme, the same is true for the other end of the sonority scale: obstruents are very exceptional, whereas vowels are obligatory in this constituent. The syllable thus can be divided into two separate "fields": the onset is the consonantal field, the rhyme the vocalic field. Apart from this essential difference, similar principles of construction apply.

The Principle of Maximal Differentiation is an example of such a principle. It goes without saying that if the first segment is a vowel and the second a consonant, these two are already sufficiently different. But we also already saw that [i] is very close to [j] and [u] to [ŭ]. The Principle of Maximal Differentiation now would hold that those two should not occur together in a rhyme.

That seems to be right: we have syllables such as *mejlo* 'mile', *homoj* 'people' and *tuj* 'immediately', but there are no Esperanto words with *[ij]. Similarly, although we have *aŭ* 'or', *poŭpo* 'rear', *Eŭropo* 'Europe', there are no words with *[uŭ].[41]

On the other hand, the fact that the consonants in the rhyme have to be sonorant is somewhat contrary to our Principle, for these consonants are clearly much more similar

[38] An alternative is that *aj* and *oj* are analysed as complex vowels, on a par with the complex consonants [c, ĵ,...].

[39] Cf. Van Oostendorp (1995) for examples from Dutch and French.

[40] Furthermore, the complex segment [ĉ] in *eĉ* 'even' is alveolar and therefore it may be extrasyllabic. In that case there are only two real exceptions left: *sub* and *nek*.

[41] It should be noted that I also have not been able to find [iŭ] and that [oŭ] also is infrequent at best.

to vowels than obstruents. We thus have a conflict between the Principle and the requirement in (23). In that case, apparently (23) decides (see Prince and Smolensky 1993 for a formal discussion of the idea of conflict resolution among conditions on phonological structure.) As a matter of fact, this should be true for the onset as well, because as far as the Principle of Maximal Differentiation is concerned, the best governee in an onset is a vowel; but these do not occur in that position, as we have seen.

3.3. Rhymes and onsets of syllables inside the word

I have already mentioned my assumption that the *word*, i.e., the unity of at least one morpheme plus a grammatical ending is the basis for syllabification in Esperanto. More specifically, I assume that the following holds:

Principle 7. Principle of Full Syllabification *(PFS). All segments in a word are syllabified together (except for peripheral alveolars).*

This principle can be seen as a somewhat more specific instance of Itô's prosodic licensing cited above. The PFS holds that every consonant cluster in the language should consist of a wellformed rhyme ending plus a well-formed onset. This is indeed usually the case:

(24)	po[r]	[t]abelo	pa[rt]o
	a[l]	[f]unto	go[lf]o
	e[n]	[kr]ii	a[nkr]o
	je[n]	[gv]idi	li[ngv]o

Yet there are more restrictions on word-internal consonant clusters. Most importantly, in most languages of the world, these clusters seem to obey the so-called Syllable Contact Law (SCL, Vennemann 1988):

Principle 8. Syllable Contact Law *(SCL). In a consonant cluster C_iC_j, if C_i is in a rhyme and C_j in an onset, C_i is preferably more sonorous than C_j.*

There are more restrictions. In the first place, "exceptional" onsets such as [kn, gn, kv] do not occur word-internally. There are words such as *akvo* 'water', but this could be syllabified as [ak.vo]; what is missing, is words such as [an.kvo].

Also the number of possible extrasyllabic consonants is much smaller within the word than at the edges. We have seen that before the word-inital onset, both [s] and [ŝ] can function as extrasyllabic segments. At the end of the word we may have [s, t, n, ĉ]. *Within* the word we only find the segment that belongs to the intersection of these two sets, viz. [s]. In the following table I underlined the extrasyllabic segments:[42]

(25)	tran$\underline{\text{s}}$	$\underline{\text{s}}$tari	ek$\underline{\text{s}}$ter
	cen$\underline{\text{t}}$	*$\underline{\text{t}}$pari	*ek$\underline{\text{t}}$ser
	*tran$\underline{\hat{\text{s}}}$	$\underline{\hat{\text{s}}}$tato	*ek$\underline{\hat{\text{s}}}$ter

[42]There is a set of words in which a [k] can occur in an extrasyllabic position if it is preceded by a sonorant and followed by a alveolar: *punkto* 'dot', *arkta* 'arctic'. See Van Oostendorp (1998a) for discussion.

There is another, even more surprising, fact, illustrated by the word *ekster* 'outside': the fact that [k] ends the rhyme of the first syllable. I have shown that word-finally obstruents such as [k] hardly ever occur in the rhyme. Word-medially, on the other hand, the rhyme position is filled by an obstruent quite frequently:

(26)	fa[kt]o	ado[pt]i	re[st]i
	a[kc]ento	ka[pt]i	la[st]a
	e[kz]ameno	ka[ps]ulo	ma[st]o

Note that these clusters also do not conform to the Syllable Contact Law. The restrictions are thus much stronger on the end of the word than on its beginning.

It is interesting to see that also in these cases alveolars are always involved, this time in the onset of the following syllable. This is more generally the case for problematic cases for the SCL. Also if we have a cluster of nasals, the second one is always alveolar [n]:

(27)	hi[mn]o	da[mn]i	a[mn]estio
	*hi[nm]o	*da[nm]i	*a[nm]estio

And even in clusters of liquids, the final liquid is always an alveolar [l]:

(28)	pe[rl]o	me[rl]o
	*pe[lr]o	*me[lr]o

We can generalize these observations:

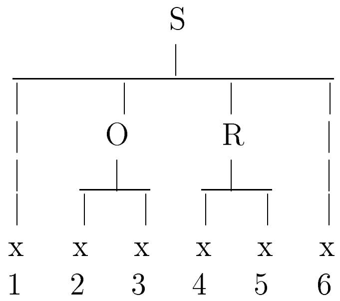
(29) **Alveolar Exception.** In clusters C_iC_j, if C_j is an alveolar, C_i does not have to conform to the sonority requirements otherwise imposed on consonants; nor does C_iC_j have to conform to the Syllable Contact Law.

The discovery of the Alveolar Exception can hardly count as a theoretical success; I leave the question of how it should be explained open for future research.

3.4. Summary

This closes our analysis of the syllable in "simple" words. We have found that the Esperanto syllable conforms to the following template,

(30)



where position 1 is occupied by alveolar fricatives, position 2 by all consonants, position 3 by consonants that can be governed by those in 2, 4 by vowels, 5 by glides or sonorant consonants (or by other consonants, if the syllable is followed by an alveolar), and position 6 by simple alveolar obstruents and possibly by [ĉ].

If a word consists of more than one syllable, we find that the consonant clusters that arise are subject to a Syllable Contact Law and an “Alveolar Exception” to that Law.

4. Morphologically complex words

In the previous section, I studied the syllable structure in the underived word. Here, I will briefly discuss some of the complications that arise if we study morphologically complex words, both as a result of derivation and as a result of compounding.[43]

4.1. Affixes

The derivational morphology of Esperanto is rather extensive: the language has both prefixes and suffixes.

4.1.1. Suffixes

Derivational suffixes always precede the grammatical endings. All suffixes, just like the grammatical endings, start with a vowel. As a matter of fact, some of them could be seen as a simple rhyme:

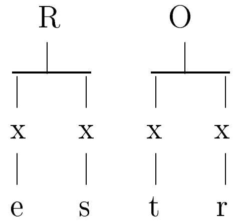
(31) *-aĉ-, -an-, -ar-, -er-, -et-, -in-, -il-, -on-, -uj-, -um-*

Yet if we look at the complete set of suffixes it becomes clear that the Esperanto derivational suffixes usually takes the form rhyme + onset (in that order):[44]

(32) *-ad-, -aĵ-, -ec-, -eg-, -estr-, -id-, -ing-, -obl-, -op-*

The most complicated suffix is *-estr-* (which is used to denote the “director” of something: *urbo* ‘city’ – *urbestro* ‘mayor’). It has the following structure:[45]

(33)



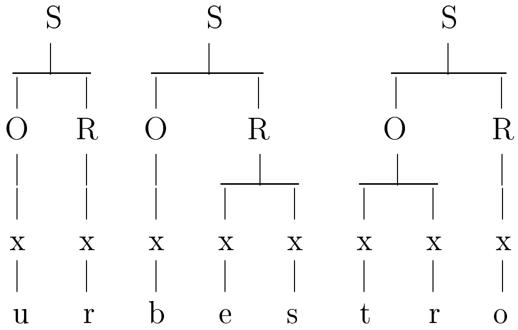
Because suffixes consist of a rhyme plus an onset, they usually can be incorporated into the syllable structure of the word quite easily: grammatical endings start with a vowel and stems usually end in something that is a possible onset:

[43] The division between derivation and compounding is not uncontroversial in the Esperanto studies literature; see Kalocsay & Waringhien (1985) and Sailer (1993), and references cited there, for discussion.

[44] I ignore the ‘learned’ suffixes *-ologi-*, etc.

[45] For the sake of simplicity, I have assumed here that [s] is in the rhyme; it could of course also be extrasyllabic.

(34)

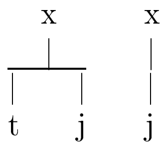


4.1.2. The exceptional suffixes *-ĉj-* and *-nj-*

Two suffixes have a markedly different structure from the others: these are *-ĉj-* and *-nj-*, forming male and female hypocoristics, respectively. These are the only suffixes not starting with a vowel; these are also suffixes that are not added to the base, but to a truncated form of the base (*patr* 'parent' > *paĉjo* 'daddy', *panjo* 'mummy').

A form such as *paĉjo* is really problematic for the theory presented here. In the first place, it has a [ĉ] in word-internal position. This [ĉ] cannot be part of the last rhyme, because complex segments do not occur in such a position. It also cannot be part of an onset, because complex segments do not govern other consonants in onsets. Furthermore the cluster [ĉj] would violate the Syllable Contact Law, while the Alveolar Exception does not usually apply to clusters ending in [j]. And finally, the sequence [ĉj] seems to contain a 'hidden long consonant'. We would represent this cluster in the following way:

(35)



Here we would thus have a double [j], not otherwise occurring in the phonology. Now it might be observed that *-ĉj-* and *-nj-* are hardly productive. Furthermore, it is hard to find minimal pairs between, e.g., *paĉjo* and hypothetical *paĉo*. Even though empirical evidence is lacking, I doubt that such pairs could exist, or that speakers could differentiate them.

Of course, we should always be careful. In 1993 a letter appeared, written by the Dutch Esperanto poet, translator and essayist Gerrit Berveling, in which it was claimed that:

> The Esperanto she [the author's daughter, who is being brought up as a native speaker] acquired, was not really Dutch-like, quite to the contrary. For example, she pronounced a word such as *paĉj'* with a clear ĉ and j, contrary to the pronunciation of speakers of Dutch.

The problem for the theory outlined here is clear: I would expect a native speaker to pronounce this word in the way that Berveling attributes to speakers of Dutch; and Dutch is my native language. On the other hand, one may wonder how Berveling's daughter learned to pronounce this word in this particular way. In the same letter, Berveling also explains that "in order to correct and extend the language use that such a small child may hear, I start reading aloud poems to her [...] She listened to the melody of the words and the sound structure of the language very carefully." In other words, an important part of the language material offered to the child should have been in a rather elevated style. It is a well-known fact that in formal style people pronounce words in a way they would not use otherwise. A pronunciation of [ĉj] may well be a "hypercorrect" spelling pronunciation. I leave this open for future research.

4.1.3. Prefixes

The structure of Esperanto prefixes is somewhat, but not particularly, more complicated than that of suffixes, as one can observe by studying the following list:

(36) *bo-, dis-, ek-, eks-, fi-, for-, ge-, mal-, mis-, pra-, re-*

Almost all of these prefixes are complete syllables, which could be independent words (and sometimes are used as such). Only [ek] and [eks] have a somewhat deviant structure: they lack an onset (which is not really a problem because an onset is missing in many other words as well) and they end in an obstruent that is not alveolar: in both cases a [k], followed by an extrasyllabic [s] in [eks]. With these prefixes one can thus easily form words that do not obey the restrictions proposed hitherto. An example is *ekparoli* 'to start to speak' (from *paroli* 'to speak'), where we have a cluster of obstruents [kp] in which the second obstruent is not an alveolar.

The structure of words derived with a prefix is thus sometimes slightly more complicated than the structure of underived words, or words derived with a suffix; this is a common fact in languages of the world.

4.2. Compounds

If we look at compounds, the number of real or apparent exceptions to the generalisations made here becomes even larger. An important cause for this is that the grammatical endings can be elided in the first part of such compounds. Take for example the world *tutmonda* 'universal' (< *tuta* 'whole', *mondo* 'world'). This word can only be syllabified in one way: [tut.mon.da]. The problem here is the first syllable, [tut] which ends in an obstruent [t].

But this is only a small problem, if we start considering words such as *konsonantgrupo* 'consonant cluster' ([kon.so.nant.gru.po] < *konsonanto* 'consonant', *grupo* 'group') with a rhyme in the third syllable that is much to large.

The source of the problem in each case is the fact that the grammatical ending, which is always a rhyme in Esperanto, is left out. The consonant that could have been in the onset of that rhyme is left behind and starts getting interpreted phonetically as closing the preceding syllable: we say [tut.mon.da] and not *[tu.tmon.da]. As far as I am able to tell, this is true even in those cases where the second word start with a vowel and where resyllabification hence would be feasible: *pacama* 'peaceloving' (< *paco* 'peace', *ama* 'loving') is syllabified [pac.a.ma], not *[pa.ca.ma].

We are thus tempted to complicate our syllabification theory for compounds as well as for prefixes. Another aspect of the linguistic structure, which has been ignored in this essay until now, could also be accounted for in this revised theory: in certain (poetic) styles of speech, one can say *esperant* in stead of *esperanto* and similarly elide all other instances of the nominal grammatical ending *-o*. The last syllable of this word could be argued to be *-ant*.

Rather than complicating the syllable template, however, I prefer to use another theory, which as a matter of fact comes close to the one proposed by Kalocsay & Waringhien (1985). These scholars write about the poetic elision in non-compounded nouns that the independence of the stem in those cases is only "apparent". The grammatical *o* does not really get lost, as far as the structure is concerned.

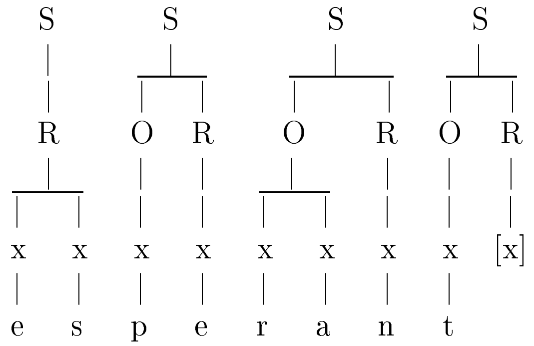
The best argument for this position is the stress system. Esperanto puts stress on the penultimate syllable of the word: *esperánto, familío*, etc. If elided vowels are really completely absent, words such as *esperánt* would mean a complication for this rule as well as for syllabification, since stress here is always on the "last" syllable. It would be simpler to say, like Kalocsay & Waringhien (1985), that "in this way, the ending *-o* remains manifestly present and recognizable in spite of the elision."

In order to express this idea of an ending which is recognizable in spite of its phonetic invisibility (and inaudibility), I want to introduce the concept **catalexis** (Kiparsky 1991, Kager 1993).

Above, I introduced the concept of extrasyllabicity: we can have segments that are not part of syllable structure. This term is in fact a special variant of extrametricality, a term from classical metric theory: in the same way in which the last or first syllable of a line can stay outside of the prosodic structure of a poem (it is not part of a foot), the last or first syllable or segment of a word can stay outside of the stress or syllable structure of the word. Now extrametricality has a counterpart in classical metrical theory: catalexis, where there is a position in the structure of a line that is not filled by any phonological material.

I propose here that Esperanto has catalectic segments in this sense: syllabic positions at the end of the word that are not filled by material. Concretely, I propose the following structure for [esperant] (I use [x] to mark a catalectic segment):

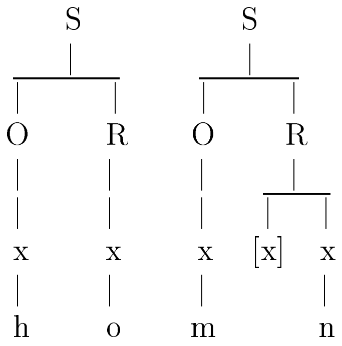
(37)



A catalectic segment thus is the exact counterpart of an extrasyllabic segment: the latter are phonetically realised but do not occupy a position, the former have a position, but no material to fill it. Extrasyllabic segments in Esperanto are always consonant, catalectic segments are always vowels (heads of a rhyme). There is no need to change the rule of stress, which can keep its simple form: stress is always on the penultimate syllable. Furthermore, there is no need to complicate the syllable structure: *esperant', bank'*, etc. and even *patr'* 'father' can be accounted for using the principles set out above.

Just like extrasyllabic segments, catalectic segments can only appear at the edge of the word. This explains why we cannot form words such as **homj* instead of *homoj* 'people', or **homn* instead of *homon* 'man (ACC)':

(38)



This catalectic segment would be non-peripheral; this is impossible. Because catalexis is possible in the first of compounds, the notion peripherality should get a somewhat more sophisticated definition: also the lefthand part of a compound counts as a word, and the vowel between two parts of the compound is in this sense peripheral.[46]

Kawasaki (1936–1953, cited by Kalocsay & Waringhien 1985) reports that in the language as it was used by Zamenhof, the ending of the first element of a compound could *not* be catalectic in the following circumstances:

1. when the consonant preceding the vowel is voiced and the consonant following it voiceless (or vice versa): *skrib[o]portanta* 'carrying writings', *kaf[o]babilo* 'kaffeeklatsch', *viv[o]fonto* 'source of life', *ŝaf[o]viro* 'sheep-man', *lud[o]tablo* 'playtable', *roz[o]kolora* 'rose coloured'

2. when the consonants preceding and following the vowel are the same: *kap[o]parto* 'part of head', *viv[o]vespero* 'evening of life', *ĉas[o]servisto* 'hunting servant'

There thus are certain independent restrictions on the application of elision: it does not happen if it would result in a cluster of consonants with different values for voicing, or if the result would be a long consonant. We have seen above that there are independent reasons to assume that these two configurations count as undesirable in Esperanto.

5. Conclusion

In this article, I studied the basic properties of the Esperanto syllable. Even though it has not been possible to solve all the problems, I believe to have pointed at least at some regularities. These regularities have in all probability not been planned by Zamenhof or other Esperanto pioneers: there are no reasons to assume that the details of phonological structure were of primary concern to them.[47]

The Esperanto syllable structure is of course very similar to that of Indo-European languages, more in particular to that of Romance and Germanic languages. This is not surprising, given the fact that most of the morphemes are borrowed from these languages. On the other hand, there is no system which has exactly the same system as Esperanto. The phonology of Italian comes close, but also this is still different. From a phonological point of view, Esperanto is an autonomous system.

[46] See Van Oostendorp (1998b) for a slightly different interpretation.

[47] Cf. Van Oostendorp (1991, 1993).

References

Albault, André

1998 La duoblaj konsonantoj en Esperanto. In: I. Haupenthal (ed.), *Menade Bal Püki Bal; Festschrift zum 50. Geburtstag von Reinhard Haupenthal.* Saarbrücken: Edition Iltis.

Berveling, Gerrit

1993 Denaske – ĉu poezie? *Esperanto* **86**, N-ro 7–8, 128–129.

Blevins, Juliette

1995 The Syllable in Phonological Theory. In: J. Goldsmith (ed.): *The Handbook of Phonological Theory.* Cambridge, Mass.: Blackwell.

Bokarev, E. A.

1966 Fonemo kaj sono en Esperanto. *Nuntempa Bulgario* **11**, vi–vii.

Chomsky, Noam

1981 *Lectures on Government and Binding.* Dordrecht: Foris.

Clements, Nick

1990 The Role of the Sonority Cycle in Core Syllabification. In: John Kingston & M. Beckman (eds.) *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech.* New York: Cambridge University Press, pp. 283–333.

Clements, Nick & S. Jay Keyser

1983 *CV Phonology.* Cambridge (Mass.): MIT Press.

Dunn, Leone

1988 Badimaya, a Western Australian Language. *Papers in Australian Linguistics.* No. 17. Canberra: Australian National University, 19–149. Pacific Linguistics.

Hayes, Bruce

1995 *Metrical Stress Theory: Principles and Case Studies.* Chicago: The University of Chicago Press.

Itô, Junko

1986 *Syllable Theory in Prosodic Phonology.* Ph.D. Dissertation, University of Massachusetts.

Itô, Junko, & Armin Mester

1993 Licensed Segments and Safe Paths. *Canadian Journal of Linguistics/Revue canadienne de linguistique* **2**, 197–213.

Kager, René

1989 *A Theory of Stress and Destressing in English and Dutch.* Dordrecht: Foris.

1995 Consequences of Catalexis. In: Harry van der Hulst & Jeroen van de Weijer (eds.), *Leiden in Last; HIL Phonology Papers I.* The Hague: Holland Academic Graphics, 269–298.

Kalocsay, Kálmán, & Gaston Waringhien

1985 *Plena Analiza Gramatiko.* Rotterdam: UEA.

Kaye, J., J. Lowenstamm & J.-R. Vergnaud

1985 The Internal Structure of Phonological Elements: A Theory of Charm and Government. *Phonology Yearbook* **2**, 305–328.

Kawasaki, N.

1936–53 *Esperantologiaj raportoj.* Osaka, Fremdlingva Universitato. 1. *Zamenhofaj ekzemploj*, 1936. 2. *Radiktipoj laŭ aranĝo de vokaloj kaj konsonantoj*, 1952. 3. *Dukonsonantaj kombinoj*, 1952. 4. *Trikonsonantaj kombinoj*, 1953.

1961 Konsonantokombinoj ĉe radikkomenco. *Esperantologio* **2**, 123–126.

Kiparsky, Paul
1991 Catalexis. Unpublished manuscript, Stanford University and Wissenschaftskolleg zu Berlin.

Levin, Juliette
1985 *A Metrical Theory of Syllabicity.* Ph.D. Dissertation, MIT.

Mangold, M.
1985 Fonemaj opozicioj en Esperanto. In: R. Haupenthal (ed.), *Li kaj ni. Festlibro por la 80a naskiĝtago de Gaston Waringhien*, Antwerpen & La Laguna, pp. 263–269.

McCarthy, John
1979 *Formal Problems in Semitic Phonology and Morphology.* Ph.D. Dissertation, MIT.

Noske, Roland
1988 La syllabification et les règles de changement de syllabe en français. In: P. Verluyten (ed.): *La phonologie du schwa français.* Amsterdam: John Benjamins, 43–88.

Okamoto, J.
1925 *Fonetika studo de Esperanto.* Tokyo: JEI.

Oostendorp, Marc van
1991 The Syntax of a Language without Grammar. MA Thesis, Tilburg University.
1993 Aangeboren principes en kunstmatige taal. *Gramma/TTT* 2.
1994 *Eseo pri la Esperanta silabo.* Unpublished manuscript.
1995 *Vowel Quality and Phonological Projection.* Ph.D. Dissertation, Tilburg University.
1998a A Note on Exceptional Syllable Structure in Esperanto. In: A. Bruyn & J. Arends, (eds.), *Mengelwerk voor Muysken; Voor Pieter Muysken bij zijn afscheid van de Universiteit van Amsterdam*, pp. 98–101. Amsterdam: UvA.
1998b Economy of Representation in the Esperanto word. In: R. van Bezooijen & R. Kager (eds.), *Linguistics in the Netherlands 1998.* Amsterdam, Benjamins, 175–187.
1999 'Leenwoordfonologie' in het Esperanto. To appear in *Gramma/TTT*.

Paradis, C. & J.-F. Prunet (eds.)
1991 *The Special Status of Coronals*, San Diego: Academic Press.

Prince, A. & P. Smolensky
1993 *Optimality Theory; Constraints Interaction and Satisfaction in Generative Grammar.* Manuscript. New Brunswick, NJ: Rutgers University, and Boulder, CO: University of Colorado.

Sailer, Manfred
1993. Complexe woorden in het Esperanto; Een generatieve herformulering van een traditionele theorie van compositie in het Esperanto. Manuscript, University of Amsterdam.

Selkirk, Elisabeth
1982 The Syllable. In: H. van der Hulst & N. Smith (eds.), *The Structure of Phonological Representations.* Vol. II: Dordrecht: Foris, 337–383.

Vennemann, Theo
1988 *Preference Laws for Syllable Structure.* Berlin: De Gruyter.

Waringhien, Gaston (ed.)
1987 *Plena Ilustrita Vortaro de Esperanto* (PIV). Reprint of the third edition, with a supplement. Paris: Sennacieca Asocio Tutmonda.

Wells, John
1978 *Lingvistikaj aspektoj de Esperanto*. Rotterdam: UEA. (Second edition 1989.)

Zamenhof, Lejzer
1927 *Lingvaj respondoj (Plena kolekto)*. Paris: Esperantista Centra Librejo. Second edition.

Author's address:
Universiteit van Amsterdam
Theoretische taalkunde
Spuistraat 210
NL-1012VT Amsterdam
E-mail address: marc.van.oostendorp@meertens.knaw.nl

Received 1999 05 01