Rhythm and semantics in the selection of linking elements

Anneke Neijt, Loes Krebbers and Paula Fikkert
University of Nijmegen/CLS & KNAW

1. Introduction

In Dutch, nominal compound formation is a very productive way of forming new words. Between two nouns that form a compound a linking element may (1a,b) or may not (1c) occur. Often the linking element is homophonous with the plural ending of the first noun, although this is not always the case, as for instance *schaapskooi in (1b) indicates, the plural of *schaap being schapen. The presence of a linking element does not always correspond to an appropriate semantic context either (i.e. referring to plurality of the left-hand constituent), nor does the absence of a linking element always correspond to the singular meaning of the left-hand constituent as shown in (1).

(1) a. linking -en- (overall 11%, after derived nouns 32.8%)
   - schapenpoot ‘leg of mutton’ (pl. schapen)
   - hoerenjong ‘child of whore’ (pl. hoeren)
   - boekenlegger ‘bookmark’ (pl. boeken)
   - rijstepap1 ‘rice porridge’ (*rijst is a sg. tantum)

b. linking -s- (overall 25%, after derived nouns 62.7%)
   - varkenspoot ‘leg of pig’ (pl. varkens)
   - schaapskooi ‘sheep fold’ (pl. schapen, *schaaps)
   - tijdsdruk ‘time pressure’ (pl. tijden, *tijds)

c. no linking element (overall 65%, after derived nouns 4.6%)
   - schaapherder ‘shepherd’ (pl. schapen)
   - burgeroorlog ‘civil war’ (pl. burgers)
   - boekwinkel ‘book shop’ (pl. boeken)

In (1) the percentages of occurrence of the -en-, -s-, and Ø linking morpheme are based on the CELEX lexical database (cf. Baayen et al. 1995, Krott et al. 2001).

Although there are a number of studies investigating the distribution of linking elements in Dutch compounds (cf. Van den Toorn 1981a,b, 1982a,b, Mattens 1987,
1990, and more recently Krott et al. 2001), this has not led to a unified account. It is not completely understood why linking elements occur in schapenpoot and varkenspoot, but not in kalkoenpoot, nor why the choice of a linking element sometimes is optional: both doelboom and doelenboom ‘target tree, goal scheme’ exist, as do tijdbesparing and tijdbesparing ‘time-saving’.² Mattens (1970, 1984) claims that linking elements are suffixes that express ‘indifferentialis’, the generic meaning of nouns, which is also used in the phrase mensen zijn sterfelijk ‘men are mortal’ or de mens is sterfelijk ‘man is mortal’. Schreuder et al. (1998) show that in writing linking (en) led to higher plurality ratings for the first constituent than (é). This suggests a closer relationship with number than Mattens’ indifferentialis would predict.

The Algemene Nederlandse Spraakkunst (Haeseryn et al. 1997:683) presents the following description:³ “Originally, the linking elements were plural endings, genitive case endings, word endings or mere transitional sounds for ease of articulation. The present situation regarding linking elements in compounds is the result of a development in which — over the centuries — genitive case was almost completely lost, words lost their endings, and several shifts occurred, particularly because of analogy effects.”

Inspired by this description, Krott et al. (2001, and to appear a,b) present a framework for accounting for the linking elements in Dutch compounds based on analogy: New compounds most likely take the linking element that occurs in existing compounds with the same left-hand constituent (either the whole constituent, its suffix, or its rime) or — although to a much lesser degree — with the same right-hand constituent. A consequence of the analogy hypothesis is that linking morphemes that reflect the language’s history — for instance, because they are the historical traces of case endings that are no longer in use — can still play a role in the minds of today’s language users and language learners. Without refuting this possibility, the present study examines the question whether other factors than analogy may also play a role, in particular the semantics of the plurality of the left-hand constituent of the compound and rhythm. Two experiments have been conducted to investigate the role of these two factors in production (both children and adults) and perception (adults only).

2. Production experiment

The first experiment aims to investigate the role of rhythm and plural semantics in production, where we focused on the choice between Ø and -en- linking morphemes.⁴ With respect to the factor rhythm, we expected that if the left-hand constituent ends in a stressed syllable and the right-hand constituent begins with a stressed syllable, the linking element -en- is favored to resolve the stress clash, as
shown in (2a),\(^3\), where a capital X stands for a stressed syllable, and a lower case x for an unstressed syllable.\(^6\) In all other contexts the insertion of the -\textit{en}- linking element will not improve the rhythmic structure of the compounds, and therefore is not predicted to occur for rhythmic reasons, as shown in (2b–c).

(2) \textit{The factor rhythm}
\begin{itemize}
  \item a. X–X \rightarrow X-\textit{en}-X
  \item b. \textit{x}x–X \rightarrow \textit{x}x–X
  \item c. X–xX \rightarrow X–xX
\end{itemize}

With respect to the factor plurality, we expected that in those cases where the left-hand constituent is semantically plural, the linking element -\textit{en}- is favored to express plural semantics. The strongest effect is expected in those cases where rhythm and plurality conspire, i.e. in cases where both rhythm and plurality favor the -\textit{en}- linking morpheme.

Furthermore, as prosody is claimed to be acquired much earlier than plurality, we expected that in children’s productions the factor rhythm would play a more important role than the factor plurality. Moreover, we expected the effects of both rhythm and plurality to be stronger in children than in adults, because the latter have more experience with data and hence will show more analogy-based behavior.

2.1 Method

Three groups of participants took part in the experiment. The first group consisted of seven children aged 4 and five children aged 5. The second group consisted of six children aged 6 and seven children aged 7. Additionally, 6 adults, all above 21 years of age, took part in the experiment.

In order to ensure that the children were capable of forming plurals with -\textit{en}, they were asked to produce the plural form of the 12 words included as left-hand members in the compound experiment. It turned out that younger children (aged 4 and 5) made some mistakes (respectively 17% and 8%), but the difference between the children was not significant (F(1,20) = 0.14, \(p<0.05\)) and all children took part in the experiment.

The test concerned an oral naming task. The participants first had to name two pictures (e.g. \textit{boom} ‘tree’, \textit{tomaat} ‘tomato’ or \textit{tomenen} ‘tomatoes’). The pictures were drawn on transparent sheets, which could be placed on top of each other. This latter situation led to the elicitation of a compound (e.g. a picture of a tree with one tomato or one with several tomatoes suggests \textit{tomaatboom} or \textit{tomenenboom}, respectively). Importantly, the compounds were not real compounds in the language, to avoid the possibility that reference was made to stored lexical representations. In this way, a set of 25 pictures for 19 concepts was used to form all 24 compound test words. Most children completed the test within half an hour. The order
Anneke Neijt, Loes Krebbers and Paula Fikkert

of the test items was determined by the participants themselves to make the task as natural as possible for them. Two children in the youngest group did not complete the task. One of them was bilingual and not capable to form compounds, the other lacked concentration. The adults finished the experiment within five minutes.

The compounds in the test set constituted four groups, as indicated in (3). If the picture denoted the plural form of the left-hand constituent of the compound (for instance a picture with several camels) the factor plurality would favor a plural form (+plural). If the two members of the compound would form a stress clash without the linking -en-, the factor rhythm would favor the linking -en- (+rhythm). The compounds indicated as (+plural, +rhythm) would thus favor the -en- linking morpheme both on semantic and on rhythmic grounds, as for instance in a tomaätenbööm (where a picture of a tree with several tomatoes was shown).

(3) a. +plural, +rhythm (both factors favor -en- responses):
tomaätenbööm, balönnenbööm, kamelenbööt, konijnenbööt, olifäntensöep, banánensöep.
b. +plural, −rhythm:
banánenballön, tomaätenballön, balönnenkonijn, blöemenkonijn, kamelengaräge, fietsergaräge.
c. −plural, +rhythm:
konijnstäart, olifäntstäart, tomaätschil, melöenschil, olifäntögen, kamelögen.
d. −plural, −rhythm (both factors disfavor -en- responses):
tomaätballön, banåanballön, olifäntgaräge, büsgaräge, përzischil, äardbeischil.

2.2 Results

The results of the production test are presented in Table 1.

With a between-items logit analysis (Rietveld and van Hout 1993, Fienberg 1980), the effect of the plural semantics, rhythm and age was measured. As predicted, there is an effect of plural semantics (F(1,65) = 46.9, p < 0.0001) and an effect of rhythm (F(1,65) = 30.9, p < 0.0001). The strongest effect is found in condition 1, where both rhythm and plural semantics favor the -en- linking element. Conversely, the weakest effect is found in condition 4 where -en- is not favored on the basis of plurality or rhythm. Yet, it is remarkable that the linking -en- is still found in a considerable amount of the data. Secondly, the semantic factor (condition 1 and 2) leads to more -en- choices than rhythm (condition 1 and 3): 94% and 91%, respectively (F(1,65) = 4.7, p = 0.03). Thirdly, age, i.e. the difference between the three groups of participants is not significant (F(2,56) = 2.4, p = 0.10).

2.3 Discussion

We hypothesized that rhythm and plural semantics would both trigger the choice
Rhythm and semantics in the selection of linking elements

Table 1. Results of the naming task. In the ±plural condition a picture of a plural (+) or singular (−) noun is presented. In the ±rhythm condition the production of the -en-linking element prevents (+) or does not prevent (−) a stress clash.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group</th>
<th>% linking -en-</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. +plural, +rhythm</td>
<td>4–5 year olds</td>
<td>96%</td>
</tr>
<tr>
<td></td>
<td>6–7 year olds</td>
<td>95%</td>
</tr>
<tr>
<td></td>
<td>adults</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>96%</td>
</tr>
<tr>
<td>2. +plural, −rhythm</td>
<td>4–5 year olds</td>
<td>95%</td>
</tr>
<tr>
<td></td>
<td>6–7 year olds</td>
<td>89%</td>
</tr>
<tr>
<td></td>
<td>adults</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>92%</td>
</tr>
<tr>
<td>3. −plural, +rhythm</td>
<td>4–5 year olds</td>
<td>85%</td>
</tr>
<tr>
<td></td>
<td>6–7 year olds</td>
<td>83%</td>
</tr>
<tr>
<td></td>
<td>adults</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>86%</td>
</tr>
<tr>
<td>4. −plural, −rhythm</td>
<td>4–5 year olds</td>
<td>57%</td>
</tr>
<tr>
<td></td>
<td>6–7 year olds</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td>adults</td>
<td>69%</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>54%</td>
</tr>
</tbody>
</table>

of linking -en- in nominal compounds. This hypothesis receives support from our experiment. We further hypothesized a larger effect of rhythm than of plurality in the behavior of children, but this turned out not to be the case. That rhythm did not have a larger effect than plural semantics may be due to the fact that these children had both acquired prosody and plural morphology, and thus used both factors in determining the linking morpheme. It is very well possible that the visual presentation of plural information has a much larger impact than the factor rhythm.

We also expected that the factors plural semantics and rhythm would have a stronger effect on children’s productions than on adults’ productions, because adults were predicted to show more analogy-based behavior due to their larger lexical experience. We anticipated children to rely more on other strategies than analogy because of their limited vocabularies. This prediction was not borne out: there is no significant difference between the age groups. A possible explanation may be, that the children that took part in the experiment are too old to show a difference, and that their vocabularies are already large enough, i.e. contained a substantial amount of compounds, for analogy to play a role.

In a post-hoc analysis we also measured the effect of left- and right-hand constituents. That is, based on CELEX (cf. Table 2) we investigated the effect of the left- and right-hand constituent in the compounds used in the experiments.
Both left and right constituents were shown to have an effect (left: \( F(1,65) = 21.5, p < 0.0001 \), right: \( F(1,65) = 3.9, p = 0.0523 \)). This leads to the conclusion that analogy as well as the factors investigated in our experiment may explain the outcome of the experiment. However, it is worth noting that the percentage of linking -en- produced in our experiment is much higher than the percentage found in general (cf. the percentages given in (1)). We will return to this point in the general discussion.

In order to test the effect of plural semantics and rhythm in the absence of a possible bias of the left- and right-hand constituent, we designed another experiment using pseudo-words.

### 3. Pseudo-word experiment

The pseudo-word experiment contained compounds formed on the basis of two pseudo-words, and consisted of a comprehension and production task. Again, the two factors considered were rhythm and plural semantics. In the production task, participants were to chose a compound with -s- or with -en- (forced choice), where we predicted that the participants would favor -en- if this would avoid a stress clash. In the perception task, the participants had to judge the plurality of the first word in the compounds with linking -en- or -s- on a seven point scale, ranging from extremely singular (1) to extremely plural (7). Here, we predicted that if -en- was not supported by rhythmic factors, i.e. was not used for clash avoidance, the
participants would be more likely to consider the first word of the compound high on plurality. For -s-, rhythm was never a factor. However, -s- is more likely to be a plural after an unstressed syllable than after a stressed syllable.

3.1 Method

38 participants, young adults (e.g. first year students of Dutch and pupils in the final year of secondary school) took part in this experiment. The words were presented orthographically. The number of pseudo-words was kept small, in order to make the task short. For syllables ending in a single consonant digraph vowels were chosen. If such syllables with digraph vowels are followed by linking -en- they do not undergo the orthographic rule of vowel letter reduction: On the basis of *gebriem* and *tip* the compounds *gebrиемtijp* and *tipengebriem* are formed. On the other hand, if a geminate vowel letter had been chosen, linking -en- would have caused vowel letter reduction: *gebraam* and *toop* would have resulted in the compounds *gebramentoop* and *topengebraam*. Unlike geminate vowels, digraph vowels show morpheme consistency in writing. We therefore chose digraph vowels to rule out the factor of morpheme consistency.

In the oral introduction of the task, most of the disyllabic pseudo-words included in the compounds were read aloud casually, in order to show their stress pattern without drawing attention on this aspect of the task. The written introduction contained no indication of the position of main stress, but the words were constructed with frequently occurring unstressed first or final syllables (often derivational prefixes and suffixes): *ge-, be-, ver-, ont-, -ik, -is, -uw, -ing*. This guaranteed that the stress pattern of the pseudo-words was predictable on the basis of the visual representation.

The first part of the test focussed on form: does a stress clash trigger the choice of -en- in pseudo-compounds? The test set contained monosyllabic and disyllabic pseudo-words which combined into 16 trisyllabic compounds, 8 with and 8 without stress clash:

(4) a. *pseudo-compounds with stress clash*
gebriemtijp, beloekpeeuw, verpeutzwink, ontmuipgarm
breunpruivik, morkbrosis, prautmeuluw, droukmoeling

b. *pseudo-compounds without stress clash*
garmontmuip, zwinkverpeut, tipgebriem, peeuwbeloek
moelingdrouk, pruivikbreun, meuluwknaaut, brosmork

The participants had to choose between two alternative realizations, either with or without linking -en-. They were presented to them on paper as shown in (5). The first line indicates the position of the word boundary within the compound:
Example of presented forms

gebriem_tijp  breun_pruivik  gar_m_ontmuip
gebriementijp breunenpruivik garmenontmuip
gebriemstijp breunspruivik garmsontmuip

In the perception part of the experiment participants had to rate the plurality of the left-hand member of pseudo-compounds on a scale ranging from 1 (singular) to 7 (plural) (cf. Schreuder et al., 1998). The following examples were given in the introduction of the task (respectively meaning ‘dog kennel’, ‘ants’ nest’, ‘boys’ bike’, and ‘ladies’ orchestra’), in order to illustrate that there are differences in plurality ratings of the first words of compounds.

The compounds with linking -en- are constructed such that for 8 forms the linking element could have been inserted for rhythm reasons and where without the linking element, a stress clash would occur. The other 8 forms contain linking -en- in a non-clash context, cf.:

We expected that the words in (7b) are considered more plural than the ones in (7a), for which -en- might occur in order to produce a better rhythm.

The compounds with -s- also consisted of two groups, modeled after the two existing patterns zinstype ‘sentence type’ and damestas ‘lady’s bag’. The linking -s- in zinstype is unlikely to be interpreted as a plural marker, since the plural form of zin is zinnen, not *zins. On the other hand, the plural form of dame is dames, which may lead to higher plurality ratings for the left-hand constituent in damestas.

Several pseudo-words from the production task of the experiment were included in the comprehension task, to ease the task for the participants. Observe that in the words constructed, -s- cannot be considered part of the onset of the following word, as this would create illegitimate onsets in Dutch: *sdr-, *sl-, etc.
3.2 Results

In the production of pseudo-compounds, -en- was chosen more often in stress clash contexts (70.7% versus 46.4%, t(37) = 4.6, p < .001) than in a context which did not have a stress clash. As predicted, the participants in the comprehension experiment considered pseudo-compounds with -en- in a stress clash context to be of lower plurality than when -en- occurred in a non-clash context (mean 4.84 versus mean 5.51; t(37) = 4.2, p < .001). For the pseudo-compounds with linking -s- the difference between the two forms presented also turned out to be significant. Where -s- could be a plural form on the basis of Dutch phonotactics, the plurality ratings were higher (mean 2.4 versus mean 2.2; t(37) = 2.1, p = 0.046, two-tailed). Note that the plurality ratings for -s- are considerably lower than the plurality ratings for -en- (mean 2.3 versus mean 5.2).

3.3 Discussion

The pseudo-word experiment confirms the conclusion of the naming task described in Section 2: both rhythm and plural semantics are relevant factors in the use of linking elements in Dutch. The two factors appear to be related in the process of comprehension: if -en- can be considered to be present for rhythmic reasons, it is considered less plural. Both -en- and -s- can function as plural morphemes, but -s- gives rise to lower plurality ratings than -en-, which is conform the findings in Neijt et al. (in preparation).

Krott (2001:226) reflects in her thesis on our pseudo-word experiment. She investigated a set of 12537 compounds gathered from the CELEX database with a left-hand constituent that takes -en as plural marker. Her results show that no linking element is the most common choice, even in stress clash situations. She finds no interaction between left and right stress, nor between right stress and linking element. In other words, she finds no support for our stress clash hypothesis in the set of existing compounds. In our experiment the participants had to make a forced choice between -en- and -s-, excluding the option of no linking element. Because this difference in experimental setting, the outcome of Krott’s investigation neither confirms nor disproves our hypothesis.

4. Conclusion

The experiments discussed show that plural semantics and rhythm are factors that influence the use and interpretation of the linking element -en- in Dutch compounds. A linking element was preferred in case of the plural semantics of the left-hand member and in a stress clash context. The interpretation of a linking element
depends on these factors as well: if the reason for a linking element is the resolution of a stress clash, the left-hand member is considered less plural.

Interestingly, no significant difference between adults and children in the production of the linking element -en- was found. Even in the youngest age group, plurality had a stronger effect than rhythm, which may be due to the fact that the pictures representing the left-hand member strongly suggested either a singular or plural response, independently of the right-hand member.

The linking -en- was the least preferred strategy in the condition [−plural, −rhythmic]. Yet, even in that condition we found a much higher percentage of -en-linking morphemes than warranted on the basis of existing compounds (cf. Krott 2001). We cannot offer an explanation for this fact. It could be the case that existing compounds follow another pattern than new formations. In this respect it seems worth investigating the historical development of linking elements. Historical changes such as the loss of case endings and the loss of allomorphy for singular nouns (e.g., both ster and sterre ‘star’ were available in earlier stages of the Dutch language) may have caused a change in the use, and ultimately the distribution of linking morphemes in compounds in Dutch, which may be reflected in the formation of new compounds by present day speakers.

Notes

* We thank Harald Baayen, Andrea Krott and Rob Schreuder for their discussions and help.

1. Observe that, although (en) is the common allomorph, linking -en- is sometimes written (e). The pronunciation of this element is usually schwa, sometimes schwa followed by a nasal, as is the case for the -en plural endings. In the present study, we focus on the -en- linking element, and will not discuss the allomorphy issue.

2. Only one rule has no exceptions: diminutives invariably take the linking -s- (pandjesjas ‘tailcoat’, meijsje ‘girl’s corpse’).

3. The translation is ours. The original text is: “De tussenklanken waren oorspronkelijk meer-voudsuitgangen, genitiefuitgangen, woorduitgangen of alleen overgangsklanken voor een gemakkelijker uitspraak. De toestand die we nu hebben bij de tussenklanken in samenstellingen is het resultaat van een eeuwenlang proces van taalontwikkeling waarin de genitief grotendeels verloren ging, de woorden hun uitgang verloren en er allerlei verschuivingen optraden, met name door de werking van de analogie.”

4. We have excluded compounds with a potential -s- linking element, because -s- does not bear on the issue of rhythm, as no extra syllable is added. Moreover, the semantics of the -s- linking morphemes can reflect either the plural or the possessive.

5. If linking elements are inserted for rhythmic reasons this may at the same time facilitate the pronunciation of these words.

6. We did not distinguish between primary and secondary stress. In Dutch nominal compounds stress falls on the main stressed syllable of the first word.
References


Toorn, M. C. van den (1981a). ‘De tussenklank in samenstellingen waarvan het eerste lid een afleiding is’. *De Nieuwe Taalgids* 74, 197–205.

