

Linking schwa in Dutch compounds: a phonomorpheme?

Anneke Neijt, Center for Language Studies, Nijmegen

In Dutch, compounds are formed with or without linking elements, cf. *zin+s+bouw* ‘sentence structure’, *woord+en+boek* ‘dictionary’ (lit. ‘word book’) and *woord+bouw* ‘word structure’. The use of linking elements has been the subject of investigations since the dissertation on Dutch compounds by van Lessen (1927), who concludes that linking elements are historic relics of stem allomorphy and case. Rule-based approaches taking a synchronic point of view (Mattens 1970, 1984 and 1987 and van den Toorn 1981 and 1982) conclude that no strict rules, but only tendencies can be formulated, and Krott (2001) shows that the combined effort of these tendencies explains only 32% of the distribution of linking elements in the compounds found in the CELEX database (Baayen, Piepenbrock and Gulikers 1995). Still, speakers of Dutch know how to use linking elements in existing compounds and although variation in the use of linking elements exists, it is not dominant; for most compounds only one form is in use. How is this knowledge available to the speaker?

Having concluded that the rule-based approach is incapable of explaining the distribution of linking elements, Krott (2001) proposes a new account based on paradigmatic analogy. In this account, linking elements in new compounds are chosen on the basis of the distribution of linking elements in existing compounds. The left constituent of a compound would be the strongest predictor of linking elements in Dutch noun-noun compounds, but also the right constituent contributes to the choice of linking elements. To give an example: on the basis of many compounds with *bank* as the left constituent and without a linking element, new forms with *bank* are predicted to occur without a linking element as well. Thus, existing forms such as *bankgeheim*, *bankgebouw*, *bankdirecteur* etc. ‘bank secret, bank building, bank manager’ predict the formation of the new forms *bankfilter*, *bankgewicht* ‘bank filter, bank weight’. Similarly, on the basis of many compounds with *boer* as the left constituent and the linking element *en*, new forms with *boer* are predicted to occur with the linking element *en* as well. Thus, existing forms such as *boerenmarkt*, *boerenschuur*, *boerenbedrijf* etc. ‘farmer’s market, farmer’s barn, farmer’s company’ predict the formation of new forms such as *boerenknobbel* and *boerenkiosk* ‘farmer’s knowledge, farmer’s stand’. Krott shows that paradigmatic analogy based on left constituents predicts 92% of the distribution of linking elements and that right constituents or semantic factors such as animacy and concreteness of the left constituent play a minor role.

In this contribution, the hypothesis is forwarded that three independent constraints play a major role in the use of linking *en* in noun-noun compounds:

- (a) *paradigmatic uniformity*, the tendency to use the same form of a word in all compounds,
- (b) *plural semantics*, the tendency to use *en* in contexts where a plural meaning of the left constituent is most appropriate, and
- (c) *rhythm*, the tendency to avoid stress clashes.

On the basis of the last two constraints, the hypothesis is forwarded that linking *en* in Dutch fulfills both a semantic and a formal function; it would be both a morpheme, and a phone, being one of the supposedly rare examples of what might be called *phonomorphemes*.

For arguments that the use of linking *en* is triggered by paradigmatic uniformity, see the thesis discussed above by Krott (2001). Arguments that plural semantics and rhythm are relevant can be found in Neijt, Krebbers and Fikkert (2002).

The hypothesis that linking schwa is used to indicate plural meaning has been forwarded in the literature mentioned (cf. also Haeseryn et al. 1997 and Schreuder et al. 1998). This hypothesis is generally accepted. But the hypothesis that rhythm is relevant has been questioned by Krott (2001:225-6), who finds no evidence for stress clash avoidance in her collection of 12537 CELEX compounds with a left constituent that takes a plural *en*. In most of the compounds with stress clash, no linking element is used. Moreover, the predicted influence of stress information of both the left and the right constituent does not increase the prediction accuracy of her simulation study with TiMBLE (Daelemans et al. 2000). Krott's conclusion is that rhythm does not reliably affect the occurrence of linking elements, at least not in existing compounds containing a left constituent with a plural *en*-form.

Here, the results of two pilot studies with new compounds will be presented that again confirm the findings in Neijt et al. (2002). Several possible explanations for the conflicting evidence found in CELEX will be forwarded in the conclusion of this paper.

The pilot studies on the choice of *en* in stress clash contexts

“In Dutch, we sometimes use *en* or *s* between the two words of a compound, and sometimes we use nothing.” With this introduction and examples of variation such as *kamelenhaar* – *kameelhaar* ‘camel hair’, *tijdverschil* – *tijdsverschil* ‘difference in time’ and *oogpotlood* – *ogenpotlood* ‘eye pencil’, 33 pupils of the final year of secondary school were instructed about variation in Dutch and asked to choose one of two variants of new or infrequently occurring compounds given in context. In four of these compounds (1a), *en* could be used to avoid a stress clash. In four other compounds with the same left constituent (1b), the right constituent of the compound begins with an unstressed syllable. In these cases, the use of *en* does not help to avoid stress clashes. When rhythm is a factor influencing the use of *en*, the participants are expected to choose *en* more often in (1a) than in (1b).

- (1) a. *New compounds where en solves stress clashes*
muntrandje / muntenrandje ‘rim of coin’
verhaalkaartje / verhalenkaartje ‘card used to help you tell a story’
gordijnwinkel / gordijnenwinkel ‘curtain shop’
landsdak / landendak ‘country roof’
- b. *New compounds where en does not solve stress clashes*
muntgevoel / muntengevoel ‘feeling for what coin is used’
verhaaldebat / verhalendebat ‘debate that also tells you a story’
gordijnfabriek / gordijnenfabriek ‘curtain factory’
landsbezit / landenbezit ‘ownership of land’

The words were embedded in a short context, such as (2):

- (2) Blinden kunnen bepalen welke munt ze gepakt hebben op grond van de vorm van de rand, op grond van het muntrandje / muntenrandje. ‘Blind people can determine what coin they have taken on the basis of the form of the rim, on the basis of the coin rim.’

Because such contexts provide a certain meaning, potential differences in interpretation presumably are reduced. The outcome of the test was as follows:

- (3) a. *66 times en where en solves stress clashes*
muntrandje 29/ muntenrandje 4
verhaalkaartje 22/ verhalenkaartje 11
gordijnwinkel 10/ gordijnenwinkel 20
landsdak 2/ landendak 31
- b. *23 times en where en does not solve stress clashes*
muntgevoel 27/ muntengevoel 6
verhaaldebat 29/ verhalendebat 4
gordijnfabriek 21/ gordijnenfabriek 12
landsbezit 32/ landenbezit 1

Observe that *en* is used more often (66 times) in stress clash contexts than in non-stress clash contexts (23 times). This illustrates that rhythm influences the choice of linking elements.

In another setting, the compounds with *mun-* and *land-* as left constituents are tested again. These left constituents never occur with linking *en* in the CELEX database. This time the participants are 13 advanced students of Dutch and the test contained more variants and the option to express ‘no preference’. The outcome again shows the influence of rhythm on the use of *en*:

- (4) a. *22 times en where en solves stress clashes*
 muntrandje 11/ munttrandje 2
 muntspelletje 5/ muntenspelletje 7/ no preference 1
 landsdak 0/ landdak 4/ landendak 8/ no preference 1
 landspeilingen 6/ landpeilingen 0/landenpeilingen 5/ no preference 2
- b. *9 times en where en does not solve stress clashes*
 muntgevoel 9/ muntengevoel 4
 muntherkenning 10/ muntetherkenning 3
 landsbezit 12/ landbezit 0/ landenbezit 1
 landsbesluiten 11/ landbesluiten 1/ landenbesluiten 1

In line with our predictions, *en* is used in a stress clash context more often (22 times) than in a non-stress clash context (9 times). The fact that the option ‘no preference’ has been used only four times shows that the participants did not have much difficulty in choosing one of the variants presented. The finding that this option occurs in (4a) exclusively might indicate that language users hesitate more often in a stress clash context. This aspect needs further investigation.

Admittedly, the number of words tested is small and a larger test should be applied to verify these findings. But the outcome of this pilot experiment is again in line with the findings of Neijt et al. (2002) and contrary the findings of Krott (2001).

Conclusions

The CELEX database presents information on the use of Dutch linking elements that differs from the outcome of our experiments with new compounds. Several explanations are available. The first one is language change: the CELEX database does not reflect our findings because rhythm is not present in earlier stages of Dutch. In the previous century, when the CELEX database was collected, only paradigmatic uniformity and plural semantics were major constraints, whereas the rhythmic constraint emerged more recently. Alternatively, it might be that existing compounds follow a different pattern because they are more conventionalized. Variation triggered by rhythm would occur more often in infrequent compounds.

The third possible explanation would refer to writing conventions. In this scenario, the systematic use of rhythm in compounds has long been present in spoken Dutch, but is not present in written Dutch because conventionally the shorter form of compounds is preferred in writing. The difference in findings then would be due to the fact that CELEX contains information on the use of written Dutch only.

The most important conclusion of the experiments with new compounds concerns the interaction of phonology and semantics. It has been shown in Neijt et al. (2002) that plural interpretation and rhythm interact in an experiment with pseudo-word compounds. The interpretation of the plurality of the left constituent depends on rhythm, such that *en* that can be interpreted as a rhythmic element leads to lower plurality ratings than *en* that cannot be interpreted as a rhythmic element. This implies that models of the internal organization of the linguistic system need to incorporate the possibility of interaction between phonology and semantics. The building blocks of words are not only phonemes and morphemes, but also a category in between would exist: the phonomorpheme.

References

- Baayen, R.H., Piepenbrock, R. and Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Haeseryn, W., Romijn, K., Geerts, G., Rooij, J. de and Toorn, M.C. van den (1997). *Algemene Nederlandse Spraakkunst*. Martinus Nijhoff Uitgevers, Groningen and Wolters Plantyn, Deurne.
- Krott, A. (2001). *Analogy in morphology. The selection of linking elements in Dutch compounds*. PhD. Dissertation, University of Nijmegen.
- Lessen, J.H. van (1927) *Samengestelde Naamwoorden in het Nederlandsch*. Groningen/Den Haag: Wolters.
- Mattens, W.H.M. (1970). *De Indifferentialis. Een onderzoek naar het anumerieke gebruik van het substantief in het algemeen bruikbaar Nederlands*. Assen.
- Mattens, W.H.M. (1984). 'De voorspelbaarheid van tussenklanken in nominale samenstellingen'. *De Nieuwe Taalgids* 77, 333–343.
- Mattens, W.H.M. (1987). 'Tussenklanken in substantivische en adjectivische samenstellingen'. *Forum der Letteren* 28, 108–114.
- Neijt, A., L. Krebbers & P. Fikkert (2002). Rhythm and semantics in the selection of linking elements.' In: H. Broekhuis & P. Fikkert (eds.) *Linguistics in the Netherlands 2002*. Amsterdam: John Benjamins.
- Schreuder, R., Neijt, A., Weide, F. van der and Baayen, R.H. (1998). 'Regular plurals in Dutch compounds: linking graphemes of morphemes?' *Language and Cognitive Processes* 13, 551–573.
- Toorn, M.C. van den (1981). 'De tussenklank in samenstellingen waarvan het eerste lid een afleiding is.' and 'De tussenklank in samenstellingen waarvan het eerste lid systematisch uitheems is.' *De Nieuwe Taalgids* 74, 197-205 and 547-552.
- Toorn, M.C. van den (1982) 'Tendenzen bij de beregeling van de verbindingsklank in nominale samenstellingen'. *De Nieuwe Taalgids* 75, 24-33 and 153-160.