Submitted

Analogical effects on linking elements in German compounds

Andrea Krott
Universtiy of Birmingham, United Kingdom

Robert Schreuder and R. Harald Baayen
University of Nijmegen, The Netherlands

Wolfgang U. Dressler
University of Vienna and the Austrian Academy of Sciences, Austria

This paper examines whether the selection of linking elements for novel German compounds can be better explained in terms of a single or a dual-route model. Previous studies had focussed on the predictability of linking elements by rules. We investigate a single-route model, by focussing on the paradigmatic analogical effect of the compounds sharing the left or right constituent with the target compound, i.e. the left (right) constituent family. A production experiment reveals an effect of the left, but not of the right constituent family. Simulation studies of the responses, using a computational model of paradigmatic analogy, show that the left constituent and its phonological and morphological properties (rime, gender, and inflectional class) simultaneously codetermine the selection of linking elements. We show how these results can be accounted for by a single-route approach, and we outline a symbolic interactive activation model that merges the factors into one psycholinguistically motivated processing mechanism.

There has been an on-going vigorous debate as to whether the processing of morphologically complex words are better accounted for by a dual-mechanism approach (e.g., Pinker & Prince, 1991; Marcus, Brinkman, Clahsen, Wiese & Pinker, 1995; Clahsen, 1999; Pinker & Ullman, 2002) or a single-mechanism approach, represented by connectionist models (e.g. Rumelhart & McClelland, 1986; Rueckl, Mikolinski, Raveh, Miner, & Mars, 1997; Plunkett & Juola, 1999). The dual-mechanism approach assumes that regular morphology is handled by rules, while irregular morphology is handled by analogy. Importantly, in terms of a double-dissociation, rule-based and analogy-based processes are assumed to have distinct neurological manifestations in the brain. In contrast, single-mechanism models deny this double-dissociation, assuming that both regular and irregular formations are the results of a single analogical mechanism. Rules are considered to be extreme forms of analogy. The debate has mainly been focussing on the English past tense formation. The dual-mechanism approach explains the formation of regular past tense forms such as *walk+ed* as the combination of the stem *walk* and the suffix *-ed*, while irregular forms such as *went* or *sang* are assumed to be stored in the mental lexicon and are retrieved as full forms. In connectionist models, however, regular and irregular formations are retrieved or built by a single mechanism, on the basis of a single neural network.

Both single and dual-route approaches do not only make different predictions as to how regular and irregular formations are processed that are already established formations of the language, but also how novel formations are created (e.g. the past tense of novel verbs that have entered the language as loanwords). In a dual-mechanism model, novel forms can either be handled by rule or analogy, although not with the same likelihood. Rule-based formations are considered to be created by a truely productive process (e.g. application of the default *-ed* rule), while formations built in analogy to stored irregular forms are considered to be rare and exceptional. In contrast, in a single-mechanism model, novel forms are always formed on the basis of analogy to existing stored forms. In other words, *-ed* is added to the novel stem in analogy to a large number of forms with *-ed*.

The rule-analogy debate has focussed very much on inflectional regularity (but see Hagiwara, Sugioka, Ito, & Kawamura, 1999; Alegre & Gordon, 1999; Clahsen, Sonnestuhl, & Blevins, 2003, for examples of derivation). In contrast, the present study examines a productive morphologi-

We are thankful for the help of Arne Fitschen and Ulrich Heid for providing us with a list of 34,000 German compounds. We also thank Kathrin Delhougne for her help in preparing these compounds for further analyses. This study was financially supported by the Dutch National Research Council NWO (PIONIER grant to the third author), the University of Nijmegen (The Netherlands), and the Max Planck Institute for Psycholinguistics (Nijmegen, The Netherlands). Requests for reprints should be addressed to Andrea Krott, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom, phone +44 (0)121 4144903, e-mail a.krott@bham.ac.uk.

cal process that underlies the formation of new words, i.e. the usage of linking elements in German compounds e.g. *-s-* in *Alter+s+Baum* 'age tree' or *-n-* in *Stelle+n+Anzeige* 'job advertisement'. We address the question whether the usage of linking elements in novel compounds is better accounted for by a single or a dual-route model. According to the dual-mechanism approach, productive processes are always rule-based. As linking elements are productively used in novel German compounds, the dual-mechanism approach predicts the existence of rules for linking elements. According to the single-mechanism approach, productive processes are based on analogy to stored existing words. Thus, a single-mechanism approach predicts that the usage of linking elements in novel formations is based on analogy only.

The occurence of linking elements between the immediate constituents of noun-noun compounds is not productive in English, but is a very common morphological phenomenon in various languages across different language families. When comparing linking elements across languages, it becomes apparent that their predictability in terms of rules varies considerably. On the one end of the scale are linking elements that only occur in frozen forms, such as the English linking *-s-* in *hunt+s+man*, *state+s+man*, *lamb+s+wool*, or *kin+s+folk*. These forms are exceptional and must therefore be stored item by item in the lexicon. On the other end of the scale are languages with linking elements that are fully predictable on the basis of rules as, for instance, Russian linking vowels. Russian root-root compounds contain *-o-* when the first root ends in a hard consonant as in *par-o-voz* (steam-O-carry 'locomotive'), otherwise they contain *-e-* as in *pyl-e-sos* (dust-E-suck 'vacuum cleaner') (Unbegaun, 1967). Such fully predictable linking elements are easily accounted for in terms of general syntagmatic rules. As a consequence, they might be generated using rules whenever a compound is produced, independent of whether this compound is already established in the language or not. For both these extreme ends of the predictability scale, the outcomes of dual and single-route models would be indistinguishable.

A more interesting group of languages lie somewhere in the middle of the predictability scale, with linking elements that are partly predictable. This appears to be typical for Germanic languages (other than English) such as Afrikaans, Danish, Dutch, Norwegian, Swedish, and German. In the case of Dutch, for example, the rules for linking elements (e.g. *-s-* and *-en-* in *tabak+s+rook* 'tabacco smell' and *schaap+en+bout* 'leg of mutton') that have been proposed in the literature (e.g., Van den Toorn, 1982a, 1982b; Haeseryn et al., 1997) do not capture all possible contexts in which linking elements can occur. Moreover, taking the subset of compounds of the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995) to which the rules are applicable, their prediction accuracy is only 63%, accounting for 32% of all Dutch compounds (Krott et al., 2001; Krott, Schreuder, & Baayen, 2002a). Not surprisingly, the search for rules In Dutch has ended with the statement that there are only tendencies and no rules (e.g. Van den Toorn, 1982a; 1982b). Recent research (Krott et al., 2001) has shown that when participants are asked to select a linking element for a novel compound, the selections can most successfully be explained on the basis of a specific form of analogy, which we will call paradigmatic analogy. In this type of analogy, the selection is based on the similarity of the target compound to a set (i.e. paradigm) of compounds, opposed to its similarity to a single exemplar, i.e. a single compound. More specifically, the selection of a linking element for a target compound has been shown to be most successfully predictable on the basis of the distribution of linking elements in the set of existing compounds that share the left constituent with the target compound. As in Krott et al., we will refer to this set of compounds as the left constituent family. For instance, the choice of the linking element for the novel compound *schaap+?+oog* ('sheep eye') is based on the distribution of linking elements in compounds such as in (1).

(1)	*schaap+en+bout*	'sheep leg'
	schaap+en+tong	'sheep tongue'
	schaap+en+wol	'lambswool'
	schaap+s+kooi	'sheep fold'
	schaap+herder	'shepherd'

In addition to the left constituent family (as the one in (1)), there is evidence for a somewhat smaller paradigmatic effect of the right constituent family, i.e. the set of compounds that share the right constituent with the target compound. In other words, the realization of the linking element in *schaap+?+oog* is co-determined by compounds such as (2), a right constituent family without clear bias for a particular linking element. Because of the stronger effect of the left constituent family and its bias for *-en-* (see (1)), *schaap+?+oog* would most probably become *schaap+en+oog*.

(2)	*varken+s+oog*	'pig eye'
	kip+en+oog	'chicken eye'
	kunst+oog	'artificial eye'

Further studies have focused on the effect of other factors on Dutch linking elements, such as the preceding suffix and the preceding rime (Krott et al., 2001; Krott, Schreuder et al., 2002a). Although these factors also appear to play a role for Dutch, they were typically overruled by the paradigmatic effect of the left constituent family.

To sum up, Dutch linking elements have been shown to be rather governed by analogy than rules. The question arises whether the outlined paradigmatic analogical account is only appropriate for Dutch linking elements, or whether it is the appropriate account for other languages with partly predictable linking elements. For that, we will focus on German because it has a much more complex system of linking elements than Dutch and because a rule-based account has been shown to be quite successful (Dressler, Libben, Stark, Pons, & Jarema, 2001; Libben, Jarema, Dressler, Stark, and Pons, 2002).

German, though etymologically close to Dutch, has maybe the most complex Germanic system of linking elements. Its main non-Latinate linking elements are *-s-*, *-e-*, *-n-*, *-en-*, *-ens-*, *-es-*, and *-er-*. In addition and in contrast to Dutch, the first constituent of a German compound may change its root vowel (via umlaut) when it is combined with a linking element (e.g., *Hand* 'hand' appears as *Händ* in the compound *Händ+e+druck* 'handshake'). It is also possible that the left constituent is shortened, i.e. reduced to its root, when it appears in a compound (e.g., *Firma* 'company' in *Firm+en+name* 'company name' or *Farbe* 'color' in *Farb+fernseher* 'color television'). The most frequent German linking element is the linking *-s-*, which occurs in 17% of all compounds in the CELEX lexical database, followed by *-(e)n-* with 15%. The remaining linking elements occur rarely(*-es-*: 1.5%; *-e-*: 1%; *-er-*: 0.4%; *-ens-*: 0.2%). Most of the noun-noun compounds, however, namely 65% of the noun-noun compounds in the CELEX lexical database, do not contain any linking elements. This is slightly less than the 69% for Dutch. As in Dutch, German linking elements have their diachronic origin in earlier inflectional forms (see Dressler and Merlini Barbaresi, 1991; Fuhrhop, 1996). This origin is still present in a number of compounds in which the left constituent together with the linking element form a possible inflected noun form. One might therefore be tempted to analyze a compound like *Wört+er+buch* 'dictionary' as word+PLURAL+book or *Himmel+s+tor* 'heaven's door' as heaven+GENITIVE+gate. However, in a lot of compounds, either the semantics of the 'suffix' is not compatible with the semantics of the compound (a *Hühn+er+ei* 'chicken(PL) egg' is not an egg produced by more than one chicken) or the combination of left constituent and linking element is not a possible inflected form (e.g., **Schwanen* in *Schwan+en+hals* 'swan neck' or **Sprach* in *Sprach+labor* 'language laboratory'). Koester, Gunter, Wagner, and Friederici (2004) have recently shown that, at least when compounds are presented auditorily, German linking elements that are equivalent to plural suffixes are not perceived as having plural semantics (but see Schreuder, Neijt, Van der Weide, & Baayen, 1998, for the plural interpretation of the Dutch linking element *-en-*). Because of the special status of German linking elements, it has even been proposed that *Hände* in *Händedruck* should not be analyzed as *Händ* followed by the linking element *-e-*, but as a single unit that serves as a compounding stem form (Fuhrhop, 1998). We will return to this issue in the general discussion.

The high number of different linking elements in German and their complex distribution seems to be related to the complex system of German noun inflection (e.g. Ortner & Müller-Bollhagen, 1991; Fuhrhop, 1996; Dressler et al., 2001). This complexity contrasts with the simpler system of noun inflection and linking elements in Dutch. Therefore the question arises whether the paradigmatic analogical approach which has been successful in accounting for Dutch (Krott et al., 2001; Krott, Krebbers, Schreuder, & Baayen, 2002; Krott, Schreuder et al., 2002a; 2002b), would also work for the much more complex situation in German.

A recent experimental study by Dressler et al. (2001) considered the question whether the choice of German linking elements is governed by rules, using a cloze-task in which participants had to create novel compounds. Dressler et al. introduced ten linguistic categories of left constituents based on grammatical gender, phonological form, and inflectional class. These categories trigger different (more or less productive) rules which insert linking elements. Thus after shwa-final feminine and animate masculine nouns a linking *-n-* is inserted productively, as in Suppe+n+topf 'soup+LINK+pot'. This productive rule competes with two unproductive rules which delete the final shwa of feminine nouns or replace it with an *-s-*, as in *Schul+∅+buch* 'school (Schule)+book' and *Geschicht+s+band* 'history (Geschichte)+LINK+volume'. Furthermore, a less productive rule inserts *-en-* after feminine and masculine consonant-final nouns, as in *Farm+en+verkauf* 'farm+LINK+sale', as well as after -a-final feminine and neuter nouns (with deletion of the word-final *-a*), as in *Firm+en+sitz* 'firm (Firma)+LINK+center'. An *-s-* is inserted productively and automatically after certain suffixes, but only optionally after consonant-final masculine and some feminine words, as in *König+s+hof* 'king+LINK+court'. The general default is, however, simple concatenation of the constituents without a linking element. Note that the participants in the present study had been Austrian German speakers. Although in general Austrian German and German German are very similar in terms of linking elements, Austrian German applies more productively (and against the default of no linking element) -n-insertion after word-final shwa of feminine nouns and -s-insertion after consonant-final masculine nouns, as in *Kohle+n+bergbau* 'coal(+LINK)+mining'.

After having distinguished these ten categories of left compound constituents, Dressler et al. determined for each category the appropriate linking element on the basis of eight (once six) exemplars. In their actual cloze task, they selected three left constituents of each category for presentation. Most of the responses were well predicted by the predefined categories and therefore support the hypothesis that German linking elements can be explained by rules. However, some categories, such as the category of root-based concatenation with truncation of the word-final shwa of a feminine noun (e.g., *Sprache* in *Sprach-labor* 'language laboratory') revealed an unexpected number of responses that deviated from the expected linking element. Dressler et al.

suggested that this variation is due to an analogical effect of the existing compounds that share the first constituent with the target compound, i.e. the left constituent family. Importantly, this category is not the only one that reveals variation. For instance, the left constituent *Stern* led to 57% *-en-* responses, which is the expected linking element, but also to 43% -∅- responses. Interestingly, 27% of the members in the constituent family of *Stern* in the CELEX contain a linking *-en-*, while 73% contain a -∅-. Even if these percentages would lead the distribution into another direction, the fact that both linking elements occur as responses again hints at an analogical effect of the left constituent family.

In a follow-up study, Libben et al. (2002) examined the speed with which novel German compounds are composed. Participants had to create and name novel compounds from two constituents presented on a computer screen. Stimuli were classified along the same ten categories as in Dressler et al. (2001). The results suggest an important role of the variability of linking elements within each category. Categories with high variability show long naming latencies, while perfectly consistent categories show short naming latencies. Libben et al.'s results resemble the ones by Krott, Schreuder, and Baayen (2002b) who found evidence for an effect of linking element variability on naming latencies when composing Dutch compounds. In contrast to Libben et al., though, Krott et al. focused on the variability within modifier families, not within left constituent categories based on inflectional classes. Together, though, both studies suggest that the distributions of linking elements in modifier-defined paradigms have a predictive power over participants' performance.

In contrast to those earlier rule-oriented studies, the aim of the present study is to investigate whether the hints for a paradigmatic analogical effect of constituent families can be confirmed in an experiment that explicitly manipulates the bias for linking elements in constituent families. Note that the idea that analogy might be involved in the formation of German compounds has already been suggested by Becker (1992). However he makes use of a very general and fuzzy notion of analogy that contrasts with the computationally tractable paradigmatic analogy with which we are concerned.

In what follows, we present a production experiment that tests the effect of both the left and the right constituent families on the three main German linking possibilities: *-s-*, *-(e)n-* , and -∅-. These three linking possibilities occur often enough in compounds to provide a substantial set of experimental items. In addition, by manipulating the bias for -∅-, we can test whether even the default choice -∅- can be explained in terms of analogy. Given the recent discussion about morphological defaults (Marcus et al., 1995; Clahsen, 1999), one would expect that default linking elements are governed by rules, not by analogy. However, if the linking elements *-s-* and *-(e)n-* are selected by analogy to their constituent families, the same might be true for -∅-. For our production experiment, we make use of the experimental design of Krott et al. (2001). Thereafter, we present simulation studies in which we predict the responses of the participants in our experiments with a computational model of analogy, TiMBL, developed by Daelemans, Zavrel, Van der Sloot, and Van den Bosch (2000). With the means of this model, we can simulate the paradigmatic effect of the left and right constituent family. In addition, we can also test whether features of the left constituent, such as rime, gender and inflectional class, for which we could not completely control for in our experiment, are better and/or additional factors influencing the selection of linking elements. Testing these features means to test the effect of general rules, similar to the ones listed in Dressler et al. (2001). In the general discussion, we outline how effects of the constituent family as well as effects of features of the left constituent such as rime or inflectional class can be modeled in a symbolic interactive activation model for analogy.

Experiment

Method

As in Krott et al. (2001), we asked participants to choose the linking elements for novel compounds. Our experimental design contained three factors: the Linking Possibility (*-s-, -en-, or* -∅-), the strength of the bias of the left constituent for that linking element (i.e. the Left Bias with the levels positive, neutral, and negative), and the strength of the bias of the right constituent for that linking element (i.e. the Right Bias with the levels positive, neutral, and negative). The three linking possibilities constitute three sub-experiments. In what follows, we will describe the materials for each sub-experiment separately.

Materials for linking possibility -s-. We determined constituent families and linking biases of these and all following experimental sets on the basis of the CELEX lexical database (Baayen et al., 1995). We constructed three sets of left constituents (L1, L2, L3) and three sets of right constituents (R1, R2, R3). The constituents of L1 and R1 had constituent families with as strong a bias as possible towards the linking *-s-*. Conversely, L3 and R3 showed a bias as strong as possible against *-s-*. The sets L2 and R2, the neutral sets, contained nouns with families without a clear preference for or against *-s-*. Each set contained 20 nouns, except for L2, for which we could only find 10 nouns.

The constituents in the L1 set had constituent family members all of which contained the linking element *-s-* in CELEX. The mean number of compounds in these families was 12.1 (range 5–46). Their mean token frequency was 402.8 per 1 million wordforms (range 0.2–1841.2). The constituents in the R1 set had CELEX constituent family members all of which also contained the linking element *-s-*. The mean number of compounds in these families was 2.3 (range 2–4). Their mean token frequency was 3 per 1 million wordforms (range 0–12.5). The neutral set L2 included left constituents whose CELEX families contained between 30% and 70% compounds with the linking element *-s-*. These families had a mean number of compounds of 3.3 (range 2–6) and a mean token frequency of 2.4 per 1 million wordforms (range 0–31.7). The constituents in the R2 set had constituent family members of which 40% to 60% contained the linking element *-s-*. These families had a mean number of compounds of 5.5 (range 3–15) and a mean token frequency of 11.5 per 1 million wordforms (range 1.2–72.8). The remaining sets L3 and R3, the groups with a bias against *-s-*, contained constituents whose family members tend not to occur with the linking *-s-* in CELEX (L3: 0%; R3: less than 20%). There were on average 2.1 (L3: range 1–9) and 2.6 (R3: range 2–6) family members, respectively, with *-s-*. Their mean token frequency was 60.5 (range 0–581.7; L3) and 4.05 (range 0–12.8; R3). Both the three left sets and the three right sets were significantly different in terms of bias for *-s-* (since assumptions for t-test were not met for these and all following comparisons, non-parametric tests were used: for all comparisons between levels, Mann-Whitney $U = 0$, two-tailed $p < .001$). These constituents were chosen to create maximal contrasts between the sets.

Materials for linking possibility -(e)n-. As for the linking *-s-*, we constructed three sets of left constituents (L1, L2, L3) and three sets of right constituents (R1, R2, R3), manipulating the bias for *-(e)n-*. Each set contained 20 nouns, except for R1, for which we could only find 18 nouns.

The properties of the sets were as follows. The constituents in the L1 set had constituent family members all of which contained the linking element *-(e)n-* in CELEX. The mean number of compounds in these families was 8.8 (range 5–22). Their mean token frequency was 927.3 per 6 million wordforms (range 0–15066). The constituents in the R1 set had constituent family members of which at least 75% contained the linking element *-(e)n-*. The mean number of compounds in these families was 2.3 (range 2–4). Their mean token frequency was 9.1 per 6 million wordforms (range 0–48). The neutral set L2 included left constituents whose families contained between 40% and 70% compounds with the linking element *-(e)n-*. These families had a mean number of compounds of 2.8 (range 2–6) and a mean token frequency of 89.0 per 6 million wordforms (range 0–707). The constituents in the R2 set had constituent family members of which 40% to 60% contained the linking element *-(e)n-*. These families had a mean number of compounds of 2.7 (range 2–7) and a mean token frequency of 12.3 per 6 million wordforms (range 0–55). The remaining sets L3 and R3, the groups with a bias against *-(e)n-*, contained constituents whose family members tend not to occur with the linking *-(e)n-* (L3: less than 5%; R3: less than 15%). There were in the mean 0.1 (L3: range 0–2) and 2.9 (R3: range 2–6) family members with *-(e)n-* respectively. Their mean token frequency was 2.7 (range 0–54; L3) and 17.3 (range 0–60; R3). Both the three left sets and the three right sets were significantly different in terms of bias for *-(e)n-* (for all comparisons between levels, Mann-Whitney $U = 0$, two-tailed $p < .001$).

Materials for linking possibility -∅-. As for the linking *-s-* and *-(e)n-*, we constructed three sets of left constituents (L1, L2, L3) and three sets of right constituents (R1, R2, R3), manipulating the bias for -∅-. Each set contained 20 nouns.

The constituents in the L1 set had constituent family members all of which contained the linking element -∅-. The mean number of compounds in these families was 15.9 (range 10–28). Their mean token frequency was 1471.4 per 6 million wordforms (range 35–9622). The constituents in the R1 set also had constituent family members of all which contained the linking element -∅-. The mean number of compounds in these families was 7 (range 5–16). Their mean token frequency was 118.7 per 6 million wordforms (range 13–911). Neutral left constituents are rare. The neutral set L2 included left constituents whose families contained between 30% and 70% compounds with the linking element -∅-. These families had a mean number of compounds of 3.3 (range 3–6) and a mean token frequency of 8757.6 per 6 million wordforms (range 0–12203). The constituents in the R2 set had constituent family members of which 30% to 70% contained the linking element -∅-. These families had a mean number of compounds of 7.6 (range 5–15) and a mean token frequency of 104.4 per 6 million wordforms (range 13–579). The remaining sets L3 and R3, the groups with a bias against -∅-, contained constituents whose family members tend not to occur with the linking -∅- (L3: less than 15%; R3: less than 20%). There were in the mean 0.4 (L3: range 0–4) and 0.1 (R3: range 0–1) family members with -∅- respectively. Their mean token frequency was 146.9 (range 0–1757; L3) and 0.4 (range 0–4; R3). Both the three left sets and the three right sets were significantly different in terms of bias for -∅- (for all comparisons between levels, Mann-Whitney $U = 0$, two-tailed $p < .001$).

As in experiments 1 and 2 in Krott et al. (2001), for each sub-experiment, each of the three sets of left constituents (L1, L2, L3) was combined with the three sets of right constituents (R1, R2, R3) to form pairs of constituents for new compounds. None of these compounds is attested in the CELEX lexical database. All are easily interpretable. Appendices A, B, and C list the experimental items of the three sub-experiments (150+174+180 = 504 items). The three item lists were combined into one experimental list, and each participant saw the list in a separate randomized order.

The results of Dressler et al. (2001) suggest that constituent families might not be the only factors that affect the choice of linking elements in novel German compound words, but that, for instance, the gender or rime of the left constituent might be important. We were not able to always control for these possibly confounding factors. For example, many items in L1 (L2) with a positive (neutral) bias for *-(e)n-* are feminine nouns ending in shwa, while L3 contains only one noun of this class. We will have to take this fact into account when interpreting the experimental results. In post-hoc analyses of our results and in the simulation studies that will follow the discussion of the experiment, we will explicitly test for confounding factors.

Procedure. As in Krott et al. (2001), the participants performed a cloze-task. The experimental list of items was presented to the participants in written form. Each line presented two nouns separated by two underscores (e.g. Zitrone__Ball). We asked the participants to combine these nouns into new compounds and to specify the most appropriate linking element, if any, at the position of the underscores, using their first intuitions (Zitrone_n_Ball). As already mentioned, a left constituent may change its form when it is combined with a linking element (e.g. umlaut of the stem vowel such as *Hühn* in *Huhn+Ei* > *Hühn+er+Ei*). We instructed participants to either mark those changes at the left constituent or to write down the full compound next to the noun pair. The experiment lasted approximately 25 minutes.

Participants. Thirty-three participants of an introductionary linguistics course at the University of Vienna volunteered to take part in the experiment. All were native speakers of German.

Results and discussion

Responses given by the participants were almost always possible German linking elements. Only twice did a participant respond with a letter that never occurs as a linking element in German. These responses were excluded from the analyses.

Table 1, 2, and 3 summarize the mean number of *-s-*, *-(e)n-* , and -∅- responses versus other responses for the three experimental subsets and the factors Left Bias and Right Bias, averaged over subjects. Appendix A, B, and C list the individual words together with the absolute numbers of responses for the noun pairs used in the three sub-experiments.

We conducted two omnibus logit analyses (see, e.g., Rietveld & Van Hout, 1993), using the log odds ratio of the

Table 1

Mean number of selected linking elements (maximum = 33) when varying the bias for -s- (positive, neutral, and negative) in the left and right compound position. Standard deviations between parentheses.

left position		right position: positive		neutral		negative	
positive	s	30.7	(3.8)	31.4	(2.6)	31.5	(3.1)
	not s	2.3	(3.8)	1.6	(2.6)	1.6	(3.1)
neutral	s	23.5	(7.5)	23.3	(9.5)	24.5	(7.5)
	not s	9.5	(7.5)	9.7	(9.5)	8.5	(7.5)
negative	s	12.0	(6.9)	13.8	(7.8)	14.7	(8.3)
	not s	21.0	(6.9)	19.3	(7.8)	18.3	(8.3)

Table 2

Mean number of selected linking elements when varying the bias for -(e)n- (positive, neutral, and negative) in the left and right compound position. Standard deviations between parentheses.

left position		right position: positive		neutral		negative	
positive	en	32.7	(0.7)	32.5	(0.8)	32.7	(0.5)
	not en	0.4	(0.7)	0.5	(0.8)	0.3	(0.5)
neutral	en	27.1	(7.5)	26.7	(7.7)	27.8	(8.4)
	not en	6.0	(7.5)	6.3	(7.7)	5.2	(8.4)
negative	en	6.8	(7.0)	9.0	(8.8)	8.8	(8.9)
	not en	26.3	(7.0)	24.1	(8.8)	24.2	(8.9)

Table 3

Mean number of selected linking elements when varying the bias for -∅- (positive, neutral, and negative) in the left and right compound position. Standard deviations between parentheses.

left position		right position: positive		neutral		negative	
positive	∅	26.9	(4.6)	26.7	(7.6)	29.0	(4.0)
	not ∅	6.1	(4.6)	6.3	(7.6)	4.1	(4.0)
neutral	∅	9.8	(9.2)	10.9	(10.0)	10.9	(10.0)
	not ∅	23.3	(9.2)	22.2	(10.0)	22.1	(10.0)
negative	∅	1.2	(3.5)	1.3	(3.2)	1.7	(4.0)
	not ∅	31.8	(3.5)	31.7	(3.2)	30.3	(4.0)

responses with the linking element in focus (*-s-*, *-(e)n-*, or *-∅-*, depending on the sub-experiment) versus other responses as the dependent variable. For a by-subject analysis (F1), we averaged responses for each subject, and for a by-item analysis (F2) we averaged responses for each noun pair. In both cases, we used Linking Possibility (*-s-*, *-(e)n-*, *-∅-*), Left Bias (positive, neutral, negative) and Right Bias (positive, neutral, negative) as fixed factors.

Both the by-subject and by-item omnibus analyses revealed main effects of the factors Linking Possibility, F1(2,864)=583.0, p<.001, F2(2,493)=160.1, p<.001, and Left Bias, F1(2,864)=2321.4, p<.001, F2(2,493)=547.8, p<.001, as well as a significant interaction between these two factors, F1(4,864)=42.3, p<.001, F2(4,493)=14.8, p<.001. The by-subject analysis also revealed a main effect of the Right Bias, F1(2,864)=11.2, p<.001, F2(2,493)=1.7, p>.05, without any further effects in neither analysis (F<1). The log odds ratio for a right positive bias appeared to be slightly higher than that of a neutral right bias, while the log odds ratio for a neutral right bias was slightly higher than that of a negative bias. Neither of these differences, however, was significant in pairwise comparisons, p>.05. Interestingly, the small, but significant effect of the right constituent family is in line with the finding for the Dutch linking elements *-en-* and *-s-* in two different sets of experimental items (Krott et al., 2001; Krott, Krebbers et al., 2002). That the effect of the right bias is only significant in the by-subject experiment in the present study suggests that its role in German is even smaller and that its importance depends very much on the compound.

Given the interaction between Left Bias and Linking Possibility, we analyzed the responses in the three sub-experiments separately, using a Bonferroni adjustment of the α-level (.017). Figure 1 gives an overview of the responses in the three sub-experiments. Both a by-subject and a by-item logit analysis show a main effect of the Left Bias on the log odds ratios for all three sub-experiments, *-s-*: F1(2,294)=372.6, p<.001, F2(2,147)=128.3, p<.001; *-(e)n-*: F1(2,294)=1781.6, p<.001, F2(2,147)=211.7, p<.001 ; *-∅-* : F1(2,294)=878.0, p<.001, F2(2,147)=224.8, p<.001. To further examine differences between the sub-experiments, we conducted pair-wise comparisons of the three bias types within each sub-experiment. Table 4 summarizes the results of these comparisons. In each experiment, a positive left bias led to higher log odds ratios than a neutral or negative bias, while a neutral bias led to higher log odds ratios than a negative bias, confirming the effect of the left bias for all three linking possibilities. Therefore, our hypothesis that the analogical effect of the left constituent family is not only relevant for *-(e)n-* and *-s-*, but also for the *-∅-* has been confirmed. In other words, even the default compounding formation is, at least in part, analogically determined.

Having seen that the left bias has very similar effects on the three linking possibilities, we further tested whether the log odds ratios of the three linking possibilities differed within each of the three levels of the left bias (positive, neutral, and negative). Table 5 summarizes the results of pairwise comparisons of the linking possibilities within each bias level. All but one comparison, namely the difference between the log odds ratio of *-s-* versus *-(e)n-* responses in the neutral bias condition are significantly different. The results show that a positive and neutral left bias led to higher log odds ratios for *-(e)n-* responses than for *-s-* responses. This order was swapped for a negative bias, as the log odds ratios for *-s-* responses were higher than the ones for *-(e)n-* responses. These differences for *-s-* and *-(e)n-* are likely to be due to small differences in bias strength, which are not apparent in the constituent families listed in CELEX. CELEX does not provide an exhaustive list of German compounds and therefore is likely not to capture differences at such a level of precision. The log odds ratios of *-∅-* responses were smaller than those of *-s-* and *-(e)n-* for all three bias levels. Thus, it seems that overall *-(e)n-* and *-s-* were more likely to be chosen as linking elements than *-∅-*. Interestingly, this result is in line with an earlier finding for Dutch linking elements. Krott et al. (2002a) report that a bias for *-∅-* can be violated in Dutch compounds more easily than a bias for *-en-* or *-s-*. This might be interpreted as a common tendency for using overt linking elements despite a majority of existing compounds without overt linking elements.

As mentioned before, in this experiment, we were not able to completely control for possible other factors that might affect the selection of German linking elements such as the rime of the left constituent. We therefore added Rime (75 different rimes), Gender (masculine vs. feminine), and In-

Table 4

Pair-wise comparisons of the effects of the three left biases (positive=pos, neutral=neu, negative=neg) on the choice of linking elements in the three sub-experiments -s-, -(e)n-, and -∅-. All $p < \alpha = .05/18 = .003$.

sub-experiment	comparison	by-subject		by-item	
		df	t	df	t
-s-	pos vs. neu	196	16.6	49	6.7
-s-	pos vs. neg	190	26.4	116	16.5
-s-	neu vs. neg	190	11.4	54	5.5
-(e)n-	pos vs. neu	184	22.9	70	7.1
-(e)n-	pos vs. neg	170	50.9	71	22.8
-(e)n-	neu vs. neg	191	27.4	114	11.6
-∅-	pos vs. neu	145	22.0	43	5.2
-∅-	pos vs. neg	172	37.0	115	25.2
-∅-	neu vs. neg	182	23.7	41	5.4

Table 5

Pair-wise comparisons of the -s-, -(e)n-, and -∅- responses in the three sub-experiments for the three left biases (positive, neutral, negative). All $p < \alpha = .05/18 = .003$ *except for comparison marked 'n.s'.*

left bias	comparison	by-subject		by-item	
		df	t	df	t
positive	-s- vs. -(e)n-	169	-7.8	91	-3.7
positive	-s- vs. -∅-	179	7.0	116	5.7
positive	-(e)n- vs. -∅-	140	14.0	85	10.2
neutral	-s- vs. -(e)n-	191	-6.2 (n.s.)	72	-2.4
neutral	-s- vs. -∅-	175	18.3	72	7.2
neutral	-(e)n- vs. -∅-	189	27.8	116	10.0
negative	-s- vs. -(e)n-	189	6.3	104	4.5
negative	-s- vs. -∅-	185	21.3	115	14.1
negative	-(e)n- vs. -∅-	195	16.6	95	6.6

flectional Class (12 classes, all provided by CELEX) of the left constituent as co-variates to the by-item analyses of the three sub-experiments[1]. We chose these co-variates because they determine the ten inflectional categories that Dressler et al. (2001) distinguished. For all three sub-experiments, there was a main effect of Left Bias, *-(e)n-* F2(2,294)=562.9, *-s-* F2(2,294)=569.5, -∅- F2(2,294)=678.6, all p<.001, and Rime, *-(e)n-* F2(39,294)=13.4, *-s-* F2(39,294)=13.2, -∅- F2(39,294)=9.4, all p<.001. In addition, there was a main effect of Gender for the *-(e)n-*, F2(1,294)=11.5, and -∅- sub-experiment, F2(1,294)=24.8, while there was a main effect of Inflectional Class in the *-s-*, F2(4,294)=5.3, and -∅- sub-experiment, F2(4,294)=4.6, all p<.001. There were no other significant effects. These results suggest that properties such as rime, gender, and inflectional class indeed affected participants' selections, but that the left bias is not confounded by these factors. They also suggest that these properties, apart from rime, are of different importance for different left constituents. To further investigate the relevance of all factors, we conducted the following simulation study.

Modeling German linking elements

In Krott et al. (2001) we have shown that selected linking elements for novel Dutch compounds, as they are given by the participants in production experiments, can be modeled with a high degree of accuracy using an exemplar-based machine-learning algorithm for the modeling of analogy, TiMBL (Daelemans, Zavrel, Van der Sloot, and Van den Bosch, 2000). Exemplar-based learning models combine similarity-based reasoning with the extensive storage of exemplars in an instance database. The class of a target, i.e. its outcome, is determined by comparing the target with the exemplars in the instance base using a set of user-specified features.[2] The most similar instance or the set of the most similar instances is used as the prediction basis.

Similar simulation studies of Krott et al. (2001) revealed that the crucial analogical factor for predicting Dutch linking elements is the left constituent, which represents the left constituent family. Prediction accuracy was enhanced when semantic class information of the right constituent was included in the feature set. Addition of the second constituent to the set did not improve prediction accuracy, although production experiments revealed clear evidence for the existence of an analogical effect of the second constituent, a non-semantic effect (Krott, Krebbers et al., 2002).

The question arises whether the choice of linking elements in German novel compounds can also be predicted by the left constituent family within such an exemplar-based modeling technique. Dressler et al. (2001) report that German linking elements are selected on the basis of ten categories of left constituents, which they interpret as evidence for rules. However, they also mention some evidence suggesting a role for analogical effects of constituent families. Simulation studies with TiMBL allow us to test whether the selected linking elements can be predicted more accurately on the basis of the left constituent family or on the basis of properties of the left constituent such as phonology, gender, and inflectional class, thus indirectly testing the predictive power of rules. Simulation studies with TiMBL can therefore test whether these features were the true factors influencing the participants' behavior in our experiment, adding to the analysis of co-variance reported above.

As a baseline study, we first ascertained to what extent constituent families and properties of the left constituent predict the linking elements of existing German compounds, namely the 8331 German compounds listed in CELEX. Table 6 lists the features that we investigated, namely the left constituent (C1), the right constituent (C2), as well as rime, gender, and inflectional class (as provided by CELEX) of the left constituent. TiMBL provides for each feature a relevance weight, the information gain (IG). The information gain measures how much information the feature contributes to the classification process. It therefore provides a first estimation of the prediction relevance of a feature. The column labeled 'celex' of Table 6 lists the information gain values for the selected features, when TiMBL is trained on all 8331

[1] Testing rime, gender, and inflectional class as co-variates in an overall analysis is not possible because the responses in the three sub-experiments were categorized differently, namely as -s- responses versus others, -en- responses versus others, and -∅- responses versus others

[2] For a description of the model's similarity metrics, see Daelemans et al. (2000) and Krott et al. (2001).

Table 6
Feature sets used in the TiMBL simulations studies of all German compounds in CELEX *(celex) and the three experiments (-S-, -EN-, -∅-) as well as their Information Gain. C1: left constituent; C2: right constituent; rime: rime of C1; gender: gender of C1; inflection: inflectional class of C1.*

		Experiments		
features	celex	-S-	-EN-	-∅-
C1	1.73	.93	.80	.79
C2	.86	.48	.51	.55
rime	1.06	.35	.23	.14
gender	.24	.02	.25	.08
inflection	.52	.04	.09	.18

Table 7
Feature sets used in the TiMBL simulations studies of all German compounds in CELEX *(celex) and the three experimental subsets (-S-, -EN-, -∅-) as well as their prediction accuracies in percentage of correctly predicted linking elements. C1: left constituent; C2: right constituent; rime: rime of C1; gender: gender of C1; inflection: inflectional class of C1.*

		Experiments		
features	celex	-S-	-EN-	-∅-
C1	87.4	79.3	79.9	80.6
C1,C2	86.9	79.3	79.9	80.6
rime	79.0	50.0	82.8	76.7
rime,gender,inflection	84.0	62.0	88.5	82.2
C1,rime,gender,inflection	91.9	79.3	79.9	80.6
agreement among participants		81.8	89.1	87.4

compounds in CELEX. The left constituent, and therefore the left constituent family, has the highest information gain value (1.73), followed by the rime of the left constituent (1.06) and the right constituent (.86). Less relevant for the classification are the inflectional class (0.52) and the gender (0.24) of the left constituent.

These values differ when training for our production experiment (-s-, -en-, and -∅-). This differences arise due to different categorization procedures. For example, in case of the experimental subset for the linking *-s-*, we classified responses as either *-s-* or *not -s-*, while, for the linking *-(e)n-*, we classified them as *-(e)n-* or *not -(e)n-*. For all three linking possibilities, just as in the baseline study, the left constituent reveals the highest information gain value. In contrast to the baseline study, the experiments suggest that the right constituent is the second most relevant feature. A comparison of the three linking elements shows that gender is more important for the experimental subset manipulating *-(e)n-* than for the subsets manipulating the other two linking possibilities, while the inflectional class is more important for -∅-. The feature rime is most relevant for the subset manipulating *-s-*. On the basis of these values, we expect that the left constituent will be the strongest predictor of German linking elements in novel compounds, followed by the right constituent. The remaining features are expected to be more or less relevant depending on the subset of target compounds.

Table 7 lists the percentage of correctly predicted linking elements in the existing German compounds in CELEX as well as in the production experiment, divided into the three experimental subsets for the three linking elements *-s-*, *-(e)n-* and -∅-.[3] The prediction accuracies given in the column 'celex' are obtained by a 'leave-one-out' procedure in which each CELEX compound is predicted on the remaining compounds. The highest prediction accuracy for a single feature is obtained by using the left constituent (87.4%). This has also been the case for the prediction of linking elements in existing Dutch compounds, although there, the left constituent predicts the selection somewhat better (92.6%) (Krott, Schreuder et al., 2002a). Note that in both languages, the model did not simply select the most frequent linking possibility. Otherwise, it would have reached a prediction accuracy of only 65%, which is the percentage of German compounds that do not contain any linking element. Surprisingly, including the right constituent in the training, the feature with the second highest information gain value, does not lead to an increase, but to a slight decrease in prediction accuracy (86.9%) of German linking elements. However, this result is in line with the results of the production experiments, in which the right constituent also did not affect the selection of linking elements in the by-item analysis. The combination of features of the left constituent, i.e. rime, gender, and inflectional class, reaches a prediction accuracy of 84%, which is significantly lower than the prediction reached by the left constituent (proportions test, $p < .001$). However, taking left constituent and its properties together leads to the high accuracy score of 91.9%, which is significantly higher than that obtained on the basis of the left constituent alone (proportions test, $p < .001$). Similarly, in the case of Dutch compounds, the combination of the left constituent, the rime and the suffix of the left constituent led to a higher prediction accuracy (93.4%) than the left constituent alone (Krott,

[3] For all reported prediction accuracies, the following parameter settings were used: similarity algorithm: IB1; feature metrics = weighted overlap; features weighed by information gain values; size of best neighbor set = 1. Different settings do not change the pattern of results. For detailed information about the parameters, see Daelemans et al. (2000).

Schreuder et al., 2002a). Thus, neither the left constituent nor its characteristics alone are sufficient to predict linking elements in existing German noun-noun compounds. It appears to be that all factors are relevant simultaneously, albeit with different weights.

The simulation studies of the responses given for novel compounds in the production experiments, however, reveal a somewhat different pattern of results. In order to predict the choices in the experiments, we compared the TiMBL's predictions with the selected linking elements that were chosen by the majority of the participants. As Table 7 shows, in both the *-s-* and the -∅- sub-experiment, the majority choices are well predicted by the left constituent (*-s-*: 79.3%; -∅-: 80.6%). Including the right constituent in the feature set does not change the results. Using just the characteristics of the left constituent leads to a decrease in prediction accuracy in the *-s-* experiment (62.0%; proportions test, $p = .002$), while it leads to a slight increase in prediction accuracy in the -∅- experiment (82.2%), which is, however, not significant (proportions test, $p = .787$). Surprisingly, in contrast to the baseline study, the combination of the left constituent and its characteristics does not improve the prediction accuracy. A different pattern emerges for the *-(e)n-* experiment. Here, combining the left constituent and its characteristics also does not increase the prediction accuracy obtained by the left constituent alone (79.9%; trained on the constituent families of the experiment). However, gender, rime, and inflectional class of the left constituent reveal a significantly higher prediction accuracy (88.5%; proportions test, $p = .040$). This result is mainly due to the rime, which alone already correctly predicts 82.8%.

Summing up, in the case of existing German compounds, a combination of the left constituent and its characteristics leads to the highest prediction accuracy. In the case of the *-s-* experiment, responses were predicted well by just the left constituent. In the *-(e)n-* experiment, responses are better predicted by the set of gender, rime, and inflectional class. In the -∅- experiment, the left constituent and the set of its properties led to very similar prediction accuracies.

One might argue that the training set of 8331 German compounds is somewhat small, when compared to the 32,000 compounds in the Dutch simulation studies. We therefore included 24,000 German compounds into the training set that were extracted from two German newspaper corpora, Frankfurter Rundschau and Stuttgarter Zeitung, which contain 76 million wordforms when combined. This allowed us to examine the effect of the two constituent families in a much broader database. This increase of training data leads to a significantly higher prediction accuracy when predicting the existing compounds in CELEX on the basis of the left constituent (93.4% versus 87.4%; proportions test, $p < .001$). However, the prediction accuracies obtained with the left constituent changed only marginally and not significantly for the novel compounds used in our experiments (*-s-*: 80.0%, $p > .05$; *-(e)n-*: 78.2%, $p > .05$; -∅-: 81.7%, $p > .05$; proportions tests). As in all previous simulation studies, the right constituent did not contribute to the prediction accuracy at all. We conclude that the prediction of the sometimes idiosyncratic patterns of linking elements in existing compounds can be improved by extending the training set. However, the patterns that are relevant for predicting linking elements in novel compound are already captured by the small set of the CELEX compounds.

The bottom row of Table 7 lists for all three experiments the mean percentages of participants that chose the linking elements that were selected by the majority of the participants. In the case of the *-s-* experiment, in the mean, 81.1% of the participants agreed with the majority choice for a linking element, while the highest prediction accuracy, based on the left constituent, was 79.3%. In the *-(e)n-* experiment, 89.1% of the participants agreed with the majority choice, while the model reaches a prediction accuracy of 88.5%, if the training is based on the rime, the gender, and the inflectional class of the left constituent. The difference between the participants' agreement (87.4%) and the model's prediction (80.6%; training on left constituent) in the -∅- experiment is not significant (proportion test, $p > .05$). We therefore conclude that, taking the highest prediction accuracies for each experiment, participants and the model appear to find the task equally difficult in all experiments. The same result was found in the simulation studies of Dutch compounds in Krott et al. (2001).

We conclude that the left constituent is the strongest predictor of linking elements in German noun-noun compounds. However, depending on the class of the left constituent, characteristics such as gender, inflectional class, and, in particular, the rime either enhance the prediction or lead to a better prediction than the constituent itself. Apparently, these factors all play a role. However, their relevance seems to vary somewhat with the type of the left constituent.

General discussion

The aim of this study was to determine whether the use of linking elements in novel German compounds is better accounted for by a single or a dual-route model. As previous studies had explored the prediction by rules (Dressler et al. 2001; Libben et al., 2002), we focused on the prediction by analogy. In particular, we focused on the paradigmatic analogical effect of the constituent families on the selection of linking elements in novel German compounds. We conducted a production experiment in which participants had to select the appropriate linking elements for novel compounds. We then tested in how far an exemplar-based computational model for analogy, TiMBL (Daelemans et al., 2000), can simulate the effect of the constituent families on participants responses and whether features of the left constituent, that had provided the basis of the rule-based account, might be better predictors of the responses.

In all three production experiments, we observed a strong paradigmatic effect of the left constituent family on the selection of linking elements, just as reported in previous studies for Dutch linking elements (Krott et al., 2001; Krott, Krebbers et al., 2002). A strong bias for a particular linking element in the left constituent family led to more responses with this linking element. The small, but significant paradigmatic effect of the right constituent family that had been found for Dutch linking elements both when averaging responses over items and subjects, appears to be of even less importance in German. The results suggest that the bias of the right constituent might affect the choice of a linking element, but that this effect depends on the specific compound.

A post-hoc analysis of co-variance with rime, gender, and inflectional class of the left constituent as co-variates revealed that the bias of the left constituent family was not confounded by these features. It did suggest, though, that these features do affect participants' choices. But, similar to the bias of the right constituent family, their effect depends on the particular compound.

Simulation studies with the exemplar-based model TiMBL, addressing the prediction of linking elements in both existing compounds and the novel compounds presented in the experiments, confirmed the experimental results. The left constituent was again the strongest predictor of linking elements in German noun-noun compounds. Just as in the by-item analysis of the experimental items, the right constituent family did not contribute to a higher prediction accuracy. For existing compounds, the simulation shows that a combined feature set of left constituent and its gender, inflectional class, and, in particular, its rime leads to highest prediction accuracies. In the case of the sub-experiment manipulating the linking *-s-*, the left constituent family was the analogical factor with the highest independent prediction accuracy, which could not be enhanced any further by including other factors. The combination of rime, gender, and inflectional class (without left constituent) led to the highest prediction accuracy in the case of the experimental subset manipulating the linking *-(e)n-*. We therefore conclude that it is neither the constituent family by itself nor properties such as rime, gender, and the inflectional class alone that affect the choice of linking elements, but a combination of these factors.

Although we did not include the categories of linking elements identified by Dressler et al. (2001) in our experimental design, a post-hoc analysis of our materials shows that each sub-experiment represents predominantly one particular subset of Dressler et al.'s categories. The subset of items with different bias for *-s-* mainly contain nouns of Dressler et al.'s categories 6 and 7, i.e. sets that both prefer the linking *-s-*. The subset with a negative bias for *-s-* mainly contains items of categories 3 and 4, nouns that are typically combined with *-n-* and *-en-*. In the case of the three subsets with different biases for *-(e)n-*, all three sets mainly contain nouns of categories 2 and 4, i.e. nouns that are typically combined with *-n* and *-en-*. Interestingly, 18 out of the 20 left constituents with a negative bias for *-(e)n-* belong to categories that, according to Dressler et al., should be combined with *-(e)n-*. In the production experiment, however, only 24% of these items were responded to with *-(e)n-*. In these cases, the constituent family clearly emerges as the stronger force. This is also true for the items in the experimental subset manipulating the bias for -∅-. These nouns mainly belong to categories that are combined with *-(e)n-* and *-s-*. Despite the predictions of the categories, participants followed the bias of the constituent families and responded with -∅-. For instance, the items with a positive bias for -∅- elicited a -∅- response in 83.3% of all cases, instead of *-en-* or *-s-*, as predicted by Dressler et al.'s categories.

A comparison of the three sub-experiments revealed that a left bias for -∅- is more easily overruled than a bias for *-e-* or *-(e)n-*, a finding that had also been attested for Dutch linking elements (Krott, Schreuder et al., 2002a). This is surprising considering the fact that the -∅- is the most common and therefore the default linking element in both German and Dutch compounds. One might argue that we are dealing with a task effect because participants were presented with the stimuli in a form that invited them to fill in a linking element. Note, though, that the results of the experiments and the simulations are very similar and the latter cannot be due to a task effect. It is therefore unlikely that our finding that a bias for -∅- is more easily overruled is a mere methodological artifact.

Considering the combined results of the simulation studies and the production experiments, both in the present study and in previous studies on German linking elements (Dressler et al., 2001; Libben et al., 2002), we conclude that German linking elements are chosen on the basis of the left constituent family as well as on the basis of properties of the left constituent such as rime, gender, and inflectional class. In contrast, Dutch linking elements are selected only on the basis of the left constituent family, while the additional paradigmatic level of properties such as the inflection class is irrelevant (Krott et al., 2001; Krott et al., 2002a). This difference between the two languages can be explained when considering the overall higher importance of inflectional class and gender in German. In contrast to Dutch, German has also case, which inflectional classes have to account for. Gender is more relevant in the German article system than in the Dutch system as well as in the definition of German inflectional classes. This shows that properties of inflection are highly relevant for word formation, which is a psycholinguistic argument for the unity of morphology and against the Split Morphology model (Anderson, 1992; Perlmutter, 1988;

for further criticism see Booij, 1994; 1996).

One can construe the functional role of gender, rime, and inflection class in German as evidence for rules that function independently of any stored exemplars, as proposed by Dressler et al. (2001). The role of the constituent families, though, must be analogical in nature. If we interpret our results as evidence for rules, the question arises why the selection of the linking *-(e)n-* is better explained by rules, while the selection of *-s-* is better explained by analogy. It is most plausible to assume that rules and analogy affect the formation of a novel compound simultaneously and that their effectiveness varies for different left constituents. But how can we explain the difference in effectiveness? There are two alternatives that can both account for our finding. First, the rule for *-(e)n-* might win the race against analogy because it is more productive and presents the default among feminines in shwa. Note that it has been suggested that rules operate faster than analogy (Anshen & Aronoff, 1988; MacWhinney, 1975). Second, there might be no abstract generalizations, i.e. no rules, but rules might be an extreme form of (highly consistent) analogy. That means we can treat both the effect of the left constituent family and the effect of properties of the left constituent as being analogical in nature (as in the TiMBL simulations). This approach allows us to explain differences in effectiveness of different factors for different linking elements. Consider, for instance, two constituents C1 and C2. C1's constituent family has a bias for *-s-*, while existing compounds with the same features as C1 (rime, gender, and inflectional class) do not show any bias for any linking element. In contrast, C2's constituent family has a bias for *-s-*, while existing compounds with the same features (rime, gender, and inflectional class) as C2 prefer the linking *-n-* (e.g., feminine nouns ending in shwa). In the case of C1, only the constituent family shows a preference, and will therefore be the best predictor for the linking element in a novel compound. In the case of C2, the constituent family predicts a different linking element than the compounds that are similar in rime and gender. As the latter set of compounds is much larger than the first set, it can have a larger influence on the outcome, and the linking *-n-* will most likely be chosen for a novel compound.

We can account for these analogical processes with a psycholinguistically motivated model, as developed by Krott et al. (2002b) for Dutch linking elements. Krott et al. report a computational symbolic interactive activation model that captures the analogical effect of the constituent families on the choice of Dutch linking elements. In this model, compounds and linking elements have independent representations. The left and right constituent of a target compound activate the compounds of their constituent families, which in their turn activate their linking elements. The selection of German linking elements can be understood along similar lines. A novel compound can activate both its left constituent family and the constituent families of other left constituents that share features such as rime, gender, and inflectional class. Figure 2 illustrates the activation flow for the novel compound *Seife+?+Stift* 'soap pen'. The semantic representation of the left constituent *Seife* sends activation to the members of its constituent family on the wordform level, such as *Seife+n+Schaum* 'lather', *Seife+n+Pulver* 'soap powder', and *Seife+n+Blase* 'soap-bubble'. In addition, it also sends activation to compounds whose left constituents are feminine nouns that end in shwa, such as *Rose+n+Wasser* 'rose water' *Seide+n+Papier* 'tissue paper', *Kreide+Zeichnung* 'chalk drawing', and *Ausname+Fall* 'exceptional case'. All these compounds then propagate activation onwards to their linking elements. The linking element that receives the most activation is selected for insertion in *Seife+?+Stift*. In our experiment, it was the *-n-* that was chosen most often (94%) for this particular compound. This example shows that, even if the left constituent family has a strong bias for a linking element, *-n-* in our case, compounds sharing the rime can activate other linking elements, such as the -∅-, as well. Given that the left constituent was the strongest predictor in our simulation studies, we assume that the left constituent family passes on more activation to the linking elements than other compounds. This is represented in Figure 2 by different line types of the connections (solid arrows: high activation; dotted arrow: low activation).

The outlined model presupposes that linking elements constitute independent units in the mental lexicon. This allows the model to explain the paradigmatic effects of left constituents sharing a property such as the inflectional class. Independent support for the hypothesis that linking elements are processed as separate units is provided by a visual perception study reported in Dressler et al. (2001). They found that the (orthographic) length of linking elements positively correlated with processing complexity. Nevertheless, the strong effect of the left constituent and its properties on the selection of linking elements reveals a tight connection between the left constituent and the linking element. Note that linking elements are part of the constituent's final syllable and that they group with the left constituent in coordinational structures such as *Sonn- und Feiertage* ('sundays and holidays'). This tight link between the left constituent and the linking element can be formalized by analyzing the left constituent and its linking element as a compounding stem, as proposed by Fuhrhop (1998). We will remain agnostic with respect to the relevance of the notion of the compounding stem and restrict ourselves to observing that, if so required, our psychological model can be understood as the mechanism underlying the creation of compounding stems.

The outlined model is compatible with what is generally assumed about the representation of compound words in the mental lexicon, namely that compounds are represented as units with connections to its immediate constituents (Libben, 1998; Sandra, 1990; Zwitserlood, 1994). To our knowledge, the model presented here and the corresponding one for Dutch (Krott et al., 2001; Krott et al., 2002b) are the only models of compound representation that explicitly account for linking elements. Koester et al. (2004), who examined the auditory processing of linking elements assume that linking elements are recognized as morphemes in an early stage of comprehension. They are part of the morpho-syntactic representation of a compound that is built up while the auditory signal unfolds. This representation is discarded, though, if the compound is accessed through a full form representation. Koester et al. do not mention whether they assume that linking elements are represented at a lexical level at all, as the interpretation of their results does not depend on this issue. A lexical representation is, however, compatible with their findings. Because the proposed representations are assumed to be independent of any plural suffixes, they cannot evoke plural semantics, which is one of Koester et al.'s main finding. Further research will have to show in how far our model can account for the processing of linking elements in other tasks. First evidence for the visual domain is given by a recent study examining well-formedness decisions to linking elements in Dutch compounds (Krott, Hagoort, & Baayen, 2004).

With regard to the single versus dual-route debate, our results provide evidence that not only inflection (Rumelhart & McClelland, 1986; Plunkett & Juola, 1999), but also a productive word-formation process can be modeled successfully with the means of a single-route approach. This is apparent from the fact that TiMBL's performance was very similar to that of a typical participant in our experiment. In addition, both TiMBl and our psycholinguistically motivated model do not only successfully capture analogical effects that cannot be formulated in terms of rules, but also effects that had formerly been regarded as being rule-based. Moreover, as outlined above, our analogical approach can account for the differences in effectiveness that different factors such as the bias of constituent families or the rime of the left constituent show for different linking elements

References

Alegre, M. & Gordon, P. (1999). Rule-based versus associative processes in derivational morphology. *Brain and Language*, *68*(1-2), 347-354.

Anderson, S. R. (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Anshen, F. , & Aronoff, M. (1988). Producing morphologically complex words. *Linguistics*, *26*(4), 641–655.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Becker, T. (1992). Compounding in German. *Rivista di Linguistica*, *4*(1), 5-36.

Booij, G. (1994). Against split morphology. In G. E. Booij & J. Van Marle (Eds.), *Yearbook of Morphology 1993* (p. 27-49). Dordrecht: Kluwer.

Booij, G. (1996). Inherent versus contextual inflection and the split morphology hypothesis. In G. E. Booij & J. Van Marle (Eds.), *Yearbook of Morphology 1995* (p. 1-16). Dordrecht: Kluwer.

Clahsen, H. (1999). Lexical entries and rules of language: a multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, *22*, 991-1060.

Clahsen, H., Sonnenstuhl, I., & Blevins, J. P. (2003). Derivational morphology in the German mental lexicon: A Dual Mechanism account. In R. H. Baayen & Schreuder, R. (Ed.), *Morphological structure in language processing* (p. 125-155). Berlin: Mouton de Gruyter.

Daelemans, W., Zavrel, J., Van der Sloot, K., & Van den Bosch, A. (2000). *TiMBL: Tilburg memory based learner reference guide. Version 3.0* (Technical Report No. ILK 00-01). Computational Linguistics Tilburg University.

Dressler, W. U., Libben, G., Stark, J., Pons, C., & Jarema, G. (2001). The processing of interfixed German compounds. In G. E. Booij & J. Van Marle (Eds.), *Yearbook of Morphology 1999* (p. 185-220). Dordrecht: Kluwer.

Dressler, W. U., & Merlini Barbaresi, L. (1991). Elements of morphopragmatics. In J. Verschueren (Ed.), *Levels of linguistic adaptation. Selected papers of the international pragmatics conference, Antwerp, August 17-22, 1987* (Vol. II, p. 33-51). Amsterdam/Philadelphia: John Benjamins.

Fuhrhop, N. (1996). Fugenelemente. In E. Lang & G. Zifonun (Eds.), *Deutsch - typologisch* (p. 525-550). Berlin: de Gruyter.

Fuhrhop, N. (1998). *Grenzfälle morphologischer Einheiten (border cases of morphological units)*. Tuebingen: Stauffenburg.

Hagiwara, H., Sugioka, Y., Ito, T., Kawamura, M., & Shiota, J. (1999). Neurolinguistic evidence for rule-based nominal suffixation. *Language*, *75*(4), 739–763.

Koester, D., Gunter, Th. C., Wagner, S., & Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: the auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience*, *16*(9), 1647–1668.

Krott, A., Baayen, R. H., & Schreuder, R. (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics*, *39*(1), 51-93.

Krott, A., Hagoort, P., & Baayen, R. H. (2004). Sublexical units and supralexical combinatorics in the processing of interfixed Dutch compounds. *Language and Cognitive Processes*, *19*(3), 453-471.

Krott, A., Krebbers, L., Schreuder, R. , & Baayen, R. H. (2002). Semantic influence on linkers in Dutch noun-noun compounds. *Folia Linguistica*, *36*, 7-22.

Krott, A., Schreuder, R., & Baayen, R. H. (2002a). Analogical hierarchy: exemplar-based modeling of linkers in Dutch noun-noun compounds. In R. Skousen (Ed.), *Analogical modeling: An exemplar-based approach to language* (p. 181-206). Amsterdam: John Benjamins.

Krott, A., Schreuder, R., & Baayen, R. H. (2002b). Linking elements in Dutch noun-noun compounds: Constituent families as predictors for response latencies. *Brain and Language*, *81*(1-3), 723-735.

Levelt, W. J. M. (1989). *Speaking. from intention to articulation*. Cambridge, Mass.: The MIT Press.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language*, *61*, 30-44.

Libben, G., Jarema, G., Dressler, W., Stark, J., & Pons, C. (2002). Triangulating the effects of interfixation in the processing of German Compounds. *Folia Linguistica*, 26, 23–43.

MacWhinney, B. (1975). Rules, rote and analogy in morphological formations by Hungarian children. *Journal of Child Language*, *2*, 65-77.

Marcus, G. F., Brinkman, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*, 189-256.

McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing. Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models*. Cambridge, Mass.: MIT Press.

Ortner, L., & Müller-Bollhagen, E. (1991). *Substantivkomposita (Deutsche Wortbildung, Volume IV)*. Berlin: Mouton de Gruyter.

Perlmutter, D. M. (1988). The split morphology hypothesis. In M. Hammond & Noonan, M. (Ed.), *Theoretical Morphology* (p. 79-100). San Diego, CA: Academic Press.

Pinker, S., & Prince, A. (1991). Regular and irregular morphology and the psychological status of rules of grammar. In *Proceedings of the 1991 meeting of the berkeley linguistics society.*

Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in the Cognitive Sciences*, *6*(11), 456-462.

Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science*, *23*(4), 463-490.

Rietveld, T., & Hout, R. Van. (1993). *Statistical techniques for the study of language and language behaviour*. Berlin: Mouton de Gruyter.

Rueckl, J. G., Mikolinski, M., Raveh, M., Miner, C. S., & Mars, F. (1997). Morphological priming, fragment completion, and connectionist networks. *Journal of Memory and Language*, *36*(3), 382-405.

Sandra, D. (1990). On the representation and processing of compound words: automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology*, *42A*, 529–567.

Schreuder, R., Neijt, A., Weide, F. Van der, & Baayen, R. H. (1998). Regular plurals in Dutch compounds: linking graphemes or morphemes? *Language and Cognitive Processes*, *13*, 551-573.

Unbegaun, B. O. (1967). *Russian grammar*. Oxford: Clarendon Press.

Van den Toorn, M. C. (1982a). Tendenzen bij de beregeling van de verbindingsklank in nominale samenstellingen I (Tendencies for the regulation of linking phonemes in nominal compounds I)). *De Nieuwe Taalgids*, *75*(1), 24-33.

Van den Toorn, M. C. (1982b). Tendenzen bij de beregeling van de verbindingsklank in nominale samenstellingen II (Tendencies for the regulation of linking phonemes in nominal compounds II). *De Nieuwe Taalgids*, *75*(2), 153-160.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, *9*, 341–368.

Appendix A

Materials for the manipulation of the linking *-s-*: left constituent and right constituent (number of *s* responses, number of other responses).

L1-R1: Left Position: Positive -s- Bias; Right Position: Positive -s- Bias

Verkehr Ideal (33,0); Handel Möglichkeit (33,0); Unglück Dauer (33,0); Zeitung Verbrechen (32,1); Seemann Bilanz (33,0); Übergang Drang (32,1); Amt Entwicklung (32,1); Versuch Gefährte (33,0); Staat Votum (28,5); Geburt Korrespondent (20,13); Durchschnitt Urkunde (33,0); Volk Formular (33,0); Wolf Manöver (28,5); Weihnacht Verbrecher (33,0); Alter Ausweis (33,0); Leben Körper (33,0); Ort Radius (31,2); Ausgleich Erfahrung (33,0); Teufel Gesuch (32,1); Leiden Koeffizient (31,2)

L1-R2: Left Position: Positive -s- Bias; Right Position: Neutral -s- Bias

Volk Zustand (32,1); Durchschnitt Besuch (33,0); Verkehr Vertrag (33,0); Teufel Grad (33,0); Amt Heim (28,5); Übergang Kirche (33,0); Zeitung Beamte (32,1); Alter Summe (33,0); Ort Gesellschaft (30,3); Geburt Form (22,11); Ausgleich Grenze (32,1); Wolf Gruppe (31,2); Handel Wissenschaft (33,0); Seemann Hilfe (32,1); Unglück Leistung (33,0); Weihnacht Apparat (33,0); Staat Kraft (29,4); Versuch Freiheit (32,1); Leiden Bereich (31,2); Leben Zeugnis (33,0)

L1-R3: Left Position: Positive -s- Bias; right constituent: Negative -s- Bias

Leben Wechsel (32,1); Teufel Fest (33,0); Staat Buch (23,10); Unglück Preis (33,0); Ausgleich Musik (33,0); Alter Baum (31,2); Zeitung Sucht (32,1); Amt Bruch (29,4); Wolf Wagen (32,1); Übergang Schutz (33,0); Verkehr Industrie (33,0); Handel Karte (32,1); Geburt Bericht (18,15); Versuch Linie (31,2); Seemann Leder (33,0); Weihnacht Versicherung (33,0); Ort Spiel (28,5); Leiden Wand (31,2); Volk Druck (31,2); Durchschnitt Fahrt (33,0)

L2-R1: Left Position: Neutral -s- Bias; Right Position: Positive -s- Bias

Schwein Gefährte (17,16); Ausfall Bilanz (31,2); Gut Urkunde (29,4); Verband Formular (31,2); Mitglied Verbrechen (26,7); Himmel Drang (32,1); Kalb Ideal (29,4); Tabak Votum (14,19); Stab Entwicklung (24,9); Mord Möglichkeit (12,21)

L2-R2: Left Position: Neutral -s- Bias; Right Position: Neutral -s- Bias

Stab Zustand (25,8); Himmel Freiheit (31,2); Ausfall Summe (31,2); Verband Bereich (32,1); Schwein Heim (9,24); Tabak Apparat (6,27); Gut Hilfe (27,6); Kalb Wissenschaft (29,4); Mord Form (16,17); Mitglied Freiheit (27,6)

L2-R3: Left Position: Neutral -s- Bias; Right Position: Negative-s- Bias

Kalb Fahrt (28,5); Schwein Fest (10,23); Mitglied Wand (26,7); Stab Linie (20,13); Himmel Musik (33,0); Tabak Leder (15,18); Verband Versicherung (29,4); Gut Preis (22,11); Ausfall Karte (32,1); Mord Baum (20,13)

L3-R1: Left Position: Negative -s- Bias; Right Position: Positive -s- Bias

Abbruch Erfahrung (23,10); Nachricht Möglichkeit (5,28); Überzeit Gesuch (13,20); Großmacht Gefäfhrte (24,9); Abfall Bilanz (24,9); Auflauf Körper (13,20); Auswahl Korrespondent (17,16); Pressluft Dauer (1,32); Unterschrift Formular (12,21); Antiquität Ausweis (10,23); Absprung Urkunde (33,0); Heimat Ideal (13,20); Gewalt Radius (7,26); Austausch Verbrechen (11,22); Versand Entwicklung (16,17); Seenot Manöver (13,20); Ausruf Votum (30,3); Haftpflicht Verbrecher (6,27); Unlust Drang (12,21); Umwelt Koeffizient (11,22)

L3-R2: Left Position: -s- bias; Right Position: Neutral -s- Bias

Haftpflicht Summe (8,25); Heimat Beamte (14,19); Überzeit Grenze (10,23); Pressluft Bereich (1,32); Abbruch Vertrag (26,7); Abfall Apparat (16,17); Seenot Zustand (17,16); Nachricht Wissenschaft (7,26); Austausch Form (14,19); Auflauf Hilfe (9,24); Antiquität Zeugnis (13,20); Großmacht Freiheit (23,10); Auswahl Kirche (20,13); Versand Leistung (14,19); Gewalt Besuch (10,23); Unlust Gesellschaft (10,23); Umwelt Heim (1,32); Absprung Grad (25,8); Ausruf Kraft (29,4); Unterschrift Gruppe (8,25)

L3-R3: Left Position: Negative -s- Bias; Right Position: Negative -s- Bias

Nachricht Buch (2,31); Antiquität Schutz (7,26); Unterschrift Baum (12,21); Heimat Bruch (13,20); Großmacht Wechsel (17,16); Absprung Wagen (19,14); Pressluft Musik (2,31); Versand Leder (17,16); Abfall Versicherung (18,15); Überzeit Linie (7,26); Haftpflicht Spiel (5,28); Gewalt Industrie (6,27); Unlust Sucht (8,25); Austausch Fahrt (13,20); Umwelt Druck (2,31); Abbruch Bericht (18,15); Auswahl Wand (20,13); Seenot Karte (15,18); Ausruf Preis (26,7); Auflauf Fest (13,20)

Appendix B

Materials for the manipulation of the linking *-(e)n-*: left constituent and right constituent (number of *-(e)n-* responses, number of other responses).

L1-R1: Left Position: Positive -(e)n- Bias; Right Position: Positive -(e)n- Bias

Rose Reiter (32,1); Kette Staub (32,1); Bär Deck (33,0); Küche Lärm (33,0); Straße Rauch (33,0); Suppe Honig (33,0); Zitrone Angebot (33,0); Seite Last (33,0); Börse Heft (32,1); Seife Strauß (32,1); Hölle König (33,0); Schütze Hass (33,0); Tasche Jäger (33,0); Stange Wärter (32,1); Tanne Nest (33,0); Woche Schmaus (33,0); Tinte Kugel (33,0); Glocke Batterie (33,0)

L1-R2: Left Position: Positive -(e)n- Bias; Right Position: Neutral -(e)n- Bias

Börse Reihe (31,2); Nerv Gewebe (32,1); Küche Leben (33,0); Tinte Löffel (33,0); Seife Stift (31,2); Tanne Gebirge (33,0); Treppe Bett (33,0); Stange Material (33,0); Glocke Bier (33,0); Bär Hals (32,1); Schütze Gesang (33,0); Straße Schein (33,0); Seite Zaun (33,0); Rose Zimmer (33,0); Kette Hieb (33,0); Hölle Wald (31,2); Suppe Archiv (32,1); Woche Vater (32,1); Zitrone Salat (33,0); Tasche Spitze (33,0)

L1-R3: Left Position: Positive -(e)n- Bias; right constituent: Negative -(e)n- Bias

Zitrone Ball (33,0); Bär Tag (33,0); Tinte Zeichen (33,0); Glocke Bruch (33,0); Stange Stück (33,0); Straße Land (31,2); Seite Tuch (33,0); Küche Straße (32,1); Kette Arbeit (33,0); Rose Bank (33,0); Seife Blume (33,0); Suppe Meister (33,0); Schütze Karte (33,0); Tasche Schiff (33,0); Treppe Weg (33,0); Tanne Papier (32,1); Woche Zeit (33,0); Hölle Recht (33,0); Nerv Bild (32,1); Börse Geschichte (31,2)

L2-R1: Left Position: Neutral -(e)n- Bias; Right Position: Positive -(e)n- Bias

Herr Deck (31,2); Schanze Lärm (31,2); Sinn Strauß (1,32); Christ Reiter (15,18); Alp König (33,0); Kohle Wärter (30,3); Fels Staub (22,11); Ehre Schmaus (33,0); Aufgabe Hass (32,1); Asche Kugel (30,3); Weide Jäger (27,6); Sekunde Rauch (33,0); Schwester Heft (33,0); Rebe Last (33,0); Ohr Batterie (28,5); Eiche Nest (33,0); Schmiere Honig (22,11); Scheibe Angebot (33,0)

L2-R2: Left Position: Neutral -(e)n- Bias; Right Position: Neutral -(e)n- Bias

Schwester Gebirge (25,8); Schmiere Stift (18,14); Aufgabe Archiv (31,2); Rebe Zaun (31,2); Eiche Spitze (33,0); Ehre Vater (32,1); Sinn Schein (1,32); Asche Bett (28,5); Sekunde Leben (33,0); Ohr Hals (27,6); Fels Bier (28,5); Kohle Gewebe (28,5); Irre Wald (21,12); Herr Löffel (30,3); Weide Reihe (25,8); Scheibe Salat (32,1); Christ Gesang (17,16); Schanze Material (30,3); Achse Hieb (33,0); Alp Zimmer (30,3)

L2-R3: Left Position: Neutral -(e)n- Bias; Right Position: Negative-(e)n- Bias

Schwester Tag (30,3); Asche Ball (30,3); Kohle Straße (31,2); Aufgabe Bild (31,2); Ehre Meister (32,1); Herr Bank (32,1); Alp Zeichen (26,7); Schmiere Tuch (21,12); Sekunde Arbeit (33,0); Sinn Zeit (2,31); Scheibe Blume (33,0); Weide Recht (23,10); Christ Geschichte (21,12); Eiche Papier (32,1); Achse Stück (33,0); Schanze Karte (33,0); Irre Land (22,11); Ohr Bruch (21,12); Fels Schiff (24,9); Rebe Weg (31,2)

L3-R1: Left Position: Negative -(e)n- Bias; Right Position: Positive -(e)n- Bias

Kreide Rauch (16,17); Welt Schmaus (21,12); Bank König (26,7); Flut Jäger (23,10); Saat Angebot (8,25); Aktion Lärm (0,33); Sensation Reiter (1,32); Kultur Last (7,26); Industrie Deck (1,32); Staat Hass (8,25); Granat Staub (16,17); Aufsicht Wärter (0,33); Schicht Honig (17,16); Hochzeit Heft (0,33); Tür Kugel (10,23); Ansicht Batterie (2,31); Zeitung Nest (0,33); Fabrik Strauß (3,30)

L3-R2: Left Position: -(e)n- bias; Right Position: Neutral -(e)n- Bias

Bank Hieb (20,13); Ansicht Material (1,32); Arznei Salat (12,21); Fabrik Leben (2,31); Sensation Wald (1,32); Schicht Gewebe (15,18); Kreide Hals (24,9); Granat Schein (13,20); Aufsicht Reihe (0,33); Saat Löffel (7,26); Tür Spitze (9,24); Staat Bier (9,24); Aktion Bett (0,33); Industrie Gebirge (1,32); Kultur Gesang (3,30); Zeitung Stift (0,33); Partei Zimmer (21,12); Hochzeit Archiv (0,33); Welt Vater (26,7); Flut Zaun (15,18)

L3-R3: Left Position: Negative -(e)n- Bias; Right Position: Negative -(e)n- Bias

Aktion Land (0,33); Tür Zeichen (5,28); Schicht Papier (12,21); Arznei Tuch (9,24); Ansicht Straße (0,33); Granat Bild (8,25); Staat Arbeit (4,29); Hochzeit Geschichte (1,32); Zeitung Bank (0,33); Welt Recht (21,12); Industrie Ball (2,31); Partei Schiff (21,12); Fabrik Tag (3,30); Flut Weg (11,22); Kultur Karte (3,30); Aufsicht Zeit (1,32); Kreide Bruch (18,15); Saat Blume (4,29); Bank Stück (12,21); Sensation Meister (0,33)

Appendix C

Materials for the manipulation of the linking -∅-: left constituent and right constituent (number of -∅- responses, number of other responses).

L1-R1: Left Position: Positive -∅- Bias; Right Position: Positive -∅- Bias

Wald Bombe (29,4); Tür Eisen (28,5); Berg Läufer (31,2); Preis Säule (33,0); Zahn Gerät (32,1); Zug Gelenk (15,18); Stein Schrift (29,4); Rohr Meter (28,5); Atom Wolle (33,0); Stadt Nummer (26,7); Tier Monat (29,4); Wand Note (32,1); Herz Analyse (28,5); Fisch Wurst (27,6); Transport Wächter (26,7); Tisch Kern (29,4); Mond Wolke (30,3); Tee Flasche (33,0); Fest Beere (28,5); Öl Essen (33,0)

L1-R2: Left Position: Positive -∅- Bias; Right Position: Neutral -∅- Bias

Mond Boot (28,5); Wand Bett (31,2); Stein Spiegel (28,5); Wald Sprache (27,6); Öl Tuer (33,0); Tisch Dienst (32,1); Transport Kraft (30,3); Tür Krieg (12,21); Preis Fehler (33,0); Berg Steuer (29,4); Rohr Zustand (28,5); Fisch Geist (22,10); Tier Staat (28,5); Fest Versicherung (31,2); Stadt Artikel (25,8); Zug Arzt (6,27); Tee Schule (33,0); Herz Unterricht (14,19); Zahn Lager (32,1); Atom Raum (32,1)

L1-R3: Left Position: Positive -∅- Bias; right constituent: Negative -∅- Bias

Wand Sittich (30,3); Berg Hauptstadt (28,5); Wald Haushalt (26,7); Stadt Maßregel (25,8); Rohr Standard (25,8); Öl Geschenk (31,2); Zug Person (14,19); Fisch Hunger (28,5); Mond Kanzler (26,7); Transport Sekretär (24,9); Tür Klage (21,12); Stein Produktion (28,5); Zahn Kummer (32,1); Fest Koalition (30,3); Atom Moral (31,2); Herz Lotto (20,13); Tier Streit (28,5); Tee Bauch (33,0); Tisch Nest (27,6); Preis Verrat (31,2)

L2-R1: Left Position: Neutral -∅- Bias; Right Position: Positive -∅- Bias

Meer Meter (5,28); Lamm Wächter (26,7); Kalb Analyse (4,29); Weide Bombe (5,28); Jahr Beere (0,33); Ohr Gelenk (17,16); Rebe Gerät (4,29); Kohle Säule (1,32); Mord Essen (8,25); Ei Kern (22,11); Fels Eisen (12,21); Alp Wolke (2,31); Watt Monat (26,7); Ausnahme Note (22,11); Achse Nummer (0,33); Zorn Schrift (12,21); Arzt Wolle (26,7); Verband Wurst (1,32); Tabak Flasche (23,10); Himmel Läufer (2,31)

L2-R2: Left Position: Neutral -∅- Bias; Right Position: Neutral -∅- Bias

Kohle Krieg (5,28); Lamm Arzt (16,17); Jahr Boot (4,29); Ausnahme Unterricht (29,4); Ohr Steuer (13,20); Fels Dienst (7,26); Weide Zustand (13,20); Ei Artikel (19,14); Achse Raum (0,33); Arzt Kraft (27,6); Rebe Bett (1,32); Watt Geist (26,7); Kalb Lager (0,33); Mord Tür (10,23); Meer Sprache (3,30); Alp Schule (2,31); Zorn Spiegel (17,16); Tabak Staat (24,9); Verband Fehler (1,32); Himmel Versicherung (0,33)

L2-R3: Left Position: Neutral -∅- Bias; Right Position: Negative-∅- Bias

Lamm Hunger (18,15); Weide Streit (8,25); Kohle Verrat (6,27); Verband Koalition (1,32); Ohr Kummer (6,27); Fels Haushalt (8,25); Ei Bauch (10,23); Tabak Nest (21,12); Zorn Moral (17,16); Mord Sekretär (13,20); Meer Hauptstadt (3,30); Watt Sittich (26,7); Jahr Geschenk (1,32); Ausnahme Kanzler (26,7); Arzt Klage (25,8); Alp Lotto (1,32); Kalb Maßregel (3,30); Achse Standard (0,33); Rebe Produktion (2,31); Himmel Person (0,33)

L3-R1: Left Position: Negative -∅- Bias; Right Position: Positive -∅- Bias

Kanone Eisen (0,33); Maus Gelenk (18,15); Träne Beere (0,33); Zigarette Bombe (1,32); Suppe Analyse (1,32); Leiden Note (0,33); Geburt Wolle (1,32); Rose Flasche (0,33); Glocke Wolke (0,33); Hölle Kern (1,32); Treppe Säule (0,33); Mittag Wächter (0,33); Wolf Wurst (2,31); Rippe Nummer (0,33); Bauer Essen (2,31); Seife Gerät (0,33); Sonne Monat (0,33); Schiff Meter (4,29); Reich Läufer (3,30); Seemann Schrift (1,32)

L3-R2: Left Position: -∅- bias; Right Position: Neutral -∅- Bias

Geburt Kraft (1,32); Bauer Unterricht (4,29); Wolf Krieg (1,32); Seife Boot (0,33); Maus Sprache (14,19); Schiff Zustand (2,31); Leiden Raum (0,33); Kanone Arzt (0,33); Sonne Dienst (0,33); Treppe Bett (0,33); Rose Steuer (0,33); Hölle Geist (0,33); Seemann Staat (0,33); Mittag Schule (0,33); Suppe Artikel (0,33); Glocke Tür (0,33); Reich Lager (4,29); Träne Spiegel (0,33); Zigarette Versicherung (0,33); Rippe Fehler (0,33)

L3-R3: Left Position: Negative -∅- Bias; Right Position: Negative -∅- Bias

Glocke Produktion (0,33); Träne Sittich (0,33); Reich Hunger (1,32); Seemann Sekretär (2,31); Zigarette Haushalt (0,33); Kanone Lotto (0,33); Rose Nest (0,33); Treppe Streit (0,33); Suppe Kanzler (1,32); Geburt Geschenk (0,33); Seife Bauch (0,33); Schiff Klage (1,32); Leiden Koalition (1,32); Mittag Kummer (0,33); Rippe Standard (0,33); Wolf Verrat (0,33); Sonne Hauptstadt (0,33); Bauer Maßregel (2,31); Maus Moral (16,17); Hölle Person (0,33)

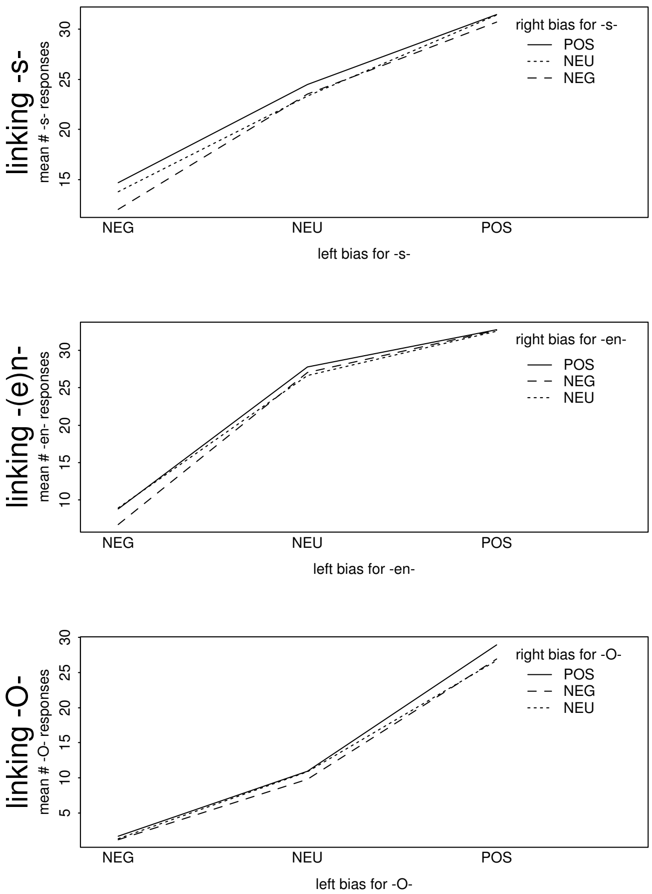

Figure 1. Interaction of biases in the left and right compound position for the linking *-s-* (upper panel), the linking *-(e)n-* (middle panel), and the linking -∅- (lower panel). POS: positive bias; NEU: neutral bias; NEG; negative bias.

Figure 1

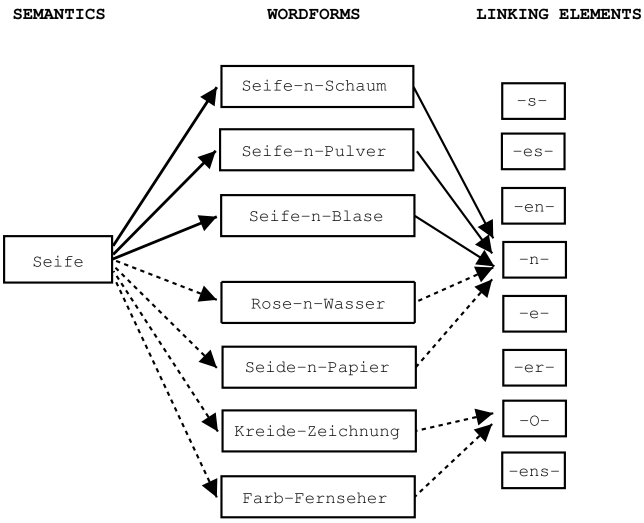

Figure 2. Connectivity in a sample part of the lexicon that is involved in the selection of the linking element for the novel German compound *Seife+?+Stift* ('soap pen'). Semantic representations (left layer); wordforms representations (lexemes in the sense of Levelt (1989), central layer) with left constituent family (upper part) and compounds sharing rime, gender, and inflectional class of the first constituent (lower part); linking elements (right layer). Line type represents amount of activation flow (solid arrow: high activation; dotted arrow: low activation).

Figure 2