Lexical representations of nouns in German rely on underspecified gender features

Psycholinguistic evidence for global underspecification

Andreas Opitz & Thomas Pechmann*

Abstract

This paper is concerned with the lexical representation of grammatical features. While by far the most of current theoretical approaches to inflectional morphology make extensive use of the two concepts of abstract feature decomposition and underspecification, psycholinguistic models of inflection, in contrast, lack such fine-grained morphological analyses almost in total. This paper reports a series of experiments that investigated the processing of grammatical gender of nouns in German. The results of these experiments support the idea that elements in the mental lexicon may be underspecified with regard to their grammatical features. But, in contrast to all established morphological and psycholinguistic approaches, we provide evidence that even the lexical representation of bare noun stems may rely on underspecified gender information. Thus, the domain of underspecification of grammatical features is extended from inflectional markers to noun stems themselves, making *underspecification a global characteristic of the mental lexicon*. We finally conclude that this fact is mainly driven by economical reasons: A feature (or a feature value) that is never used for grammatical operations (e.g., inflectional marking or evaluation of agreement) might not be necessarily represented in the language system at all.

1. Background: Syncretism and underspecification

Traditionally, instances of a certain grammatical category (e.g., gender or case) are categorically labeled to differentiate between distinct classes (e.g., masculine, feminine, neuter referring to gender in German). Current morphological theories, however, propose more fine-grained analyses of these categories. One of the most central empirical observations that lead to this view is the

* The study presented here is basically an extension of some side-findings of a PhD-Thesis which was supervised by Gereon Müller. Without his valuable and essential comments, inspirations, and critics this work wouldn't have been possible.

Topics at InfL, 235–262
A. Assmann, S. Bank, D. Georgi, T. Klein, P. Weisser & E. Zimmermann (eds.)
Linguistische Arbeits Berichte 92, Universität Leipzig 2014

	'the fork' FEMININE	'the knife' NEUTER	'the spoon' MASCULINE
NOM SG	*die Gabel*	*das Messer*	*der Löffel*
ACC SG	*die Gabel*	*das Messer*	*den Löffel*
DAT SG	*der Gabel*	*dem Messer*	*dem Löffel*
GEN SG	*der Gabel*	*des Messers*	*des Löffels*

Table 1: Example of syncretism in German nominal declension

case of inflectional syncretism in many languages of the world with fusional morphology. In such languages, an inflectional marker that represents (or realizes) one of the grammatical categories of the language (e.g., gender, case, or number) may either clearly correspond to one particular grammatical context, or it may be ambiguous by referring to more than one grammatical context. In other words, an inflectional paradigm may exhibit instances of syncretism, i.e., ambiguous morphological forms as illustrated in table 1 for inflectional marking on determiners in German.

This phenomenon of syncretic morphological marking is accounted for by *abstract feature decomposition* and the concept of *underspecification* in almost all current morphological frameworks, as in Distributed Morphology (cf. Halle and Marantz 1993, 1994), Paradigm Function Morphology (cf. Stump 2001), Minimalist Morphology (cf. Wunderlich and Fabri 1995, Wunderlich 1996), Network Morphology (cf. Corbett and Fraser 1993), various accounts in Optimality Theory (cf. Trommer 2006, Grimshaw 2001) and many others.

The over-all idea behind these two concepts is a decomposition of 'traditional' labels of morphosyntactic categories into more abstract, binary features, thereby yielding the possibility to refer to natural classes of such categories. Thus, the three instances of grammatical gender in German could be described by two abstract binary features [±fem] and [±masc]: 'feminine' [+fem, −masc], 'masculine' [−fem, +masc], 'neuter' [−fem, −masc].

Therefore, a single morphological exponent, e.g. the form of the German determiner *dem*, which occurs both in masculine and neuter contexts in dative singular, can be referred to with only one but *underspecified* set of features containing only [−fem]. With this gender specification for dative singular[1]

[1]Note that this is simplifying the actual feature specifications of the determiner. Of course,

the determiner *dem* will show up in both masculine and neuter contexts but, crucially, not in feminine contexts. This is in contrast to the alternative, traditional view that would assume two separate but homophone forms dem_1 and dem_2 both comprising of full gender specifications, i.e. [−fem, +masc] and [−fem, −masc], respectively.

2. Underspecification of grammatical features in psycholinguistics

While analyses of inflectional systems that crucially rely on abstract feature decomposition and underspecification are very common in many different frameworks in theoretical morphology (see section 1), psycholinguistic models of inflection consistently lack such fine-grained morphological analyses. Virtually all current psycholinguistic models do not make use of underspecified, abstract features specifications. This holds, e.g., for such diverse models as the Satellite Model (Lukatela et al. 1978, 1980), schema-based models (Bybee 1985, 1995), variants of connectionist models (cf. among many others Rumelhart and McClelland 1982, Seidenberg and Gonnerman 2000), serial modular models (as prominently represented by Levelt et al. 1999), the Full Decomposition Model (Stockall and Marantz 2006), the Augmented Addressed Morphology Model (Caramazza et al. 1985, 1988) and others.

Interestingly, while all these models operate with rather traditional, categorical features (implementing them as generic nodes or lexically stored features, depending on the architecture and type of the model), they are all in principle compatible with a more fine-grained featural specification. As a first step, all that would be necessary is replacing categorical features by their decomposed (and underspecified) equivalent notations.

However, it could be argued that such a 'complication' of the architecture, induced by purely theoretical considerations, would by no means be necessary in order to derive an adequate and empirically justified cognitive model of language processing. At a first glance, replacing a traditional label of, e.g., a combinatorial node like 'masculine' by [−fem, +masc] might be nothing more than a notational variant. It is not readily obvious that the retrieval or processing of, e.g., the masculine gender feature should be qualitatively different due to a different labeling. Or, to put it in other words, a relevant reason

morphological markers will have be to specified for other features such as number and case as well, which are left out here for the sake of illustration.

to implement the notions of decomposed features and underspecification into a cognitive model of language would only be given if traditional and underspecification-based approaches would make different, empirically testable predictions regarding human language processing.

This question is barely addressed in psycholinguistic research yet. Although there is first evidence that underspecification and abstract feature-decomposition are involved in human language processing (for evidence mainly based on different priming paradigms see, e.g., Penke 2006, Penke et al. 2004, Clahsen et al. 2001) their implementation into the modeling of human language processing and production has been very limited. In a recent ERP-study Opitz et al. (2013) found additional neurophysiological evidence for the cognitive reality of underspecification of morphosyntactic features. The authors investigated the processing of inflectional markers on adjectives in German phrases of the type *durch gutes Design* ('by good design') and found, using a ERP violation paradigm, that different incorrect markers on the adjective yielded different neurophysiological responses depending on which of the two different criteria (specificity versus compatibility) that are both inherently related to the notion of underspecification were violated. More precisely, the authors reported a stronger left-anterior negativity (LAN) for incompatible markers (i.e. markers that yield a feature mismatch) than for illicit markers that were in principle compatible (i.e. markers that did not yield conflicting features) but that were not specific enough. What is of importance here is that the effect was observed while the noun was processed. The authors interpret the different LAN amplitudes as mirroring the checking process of the agreement relation between adjective and noun. An example of the features involved in this checking process and the related LAN effects is given in (1)[2]. Note that only specifications for gender features are given here. Case and number features are omitted for the sake of illustration and because they do not interfere with the observed gender effects.

[2] Theoretically, there are many different abstract feature specifications possible. This is mirrored, for instance, in a number of different morphological analyses proposed in the literature (see, among others, Blevins 1995, Sauerland 1996, Wunderlich 1997, Bierwisch 1967, Gallmann 2004). However, the feature specification assumed in Opitz et al. (2013), which is basically grounded in Blevins' analysis, is chosen by a couple of conceptually reasons, e.g., it maximally reduces the number of distinct markers and it is also compatible with their empirical findings.

(1) Two types of illicit neuter agreement and related LAN effects (following Opitz et al. 2013: 242)

a. *correct*

durch	schlichtes	Design
by	plain$_{Neut}$	design$_{Neut}$
	$[-fem]$	$[-masc, -fem]$

b. *incorrect 1*

compatible (excluded by Specificity) →LAN

*durch	schlichte	Design
by	plain$_{Fem}$	design$_{Neut}$
	[ ]	$[-masc, -fem]$

c. *incorrect 2*

incompatible (excluded by Compatibility)[3] →stronger LAN

*durch	schlichten	Design
by	plain$_{Masc}$	design$_{Neut}$
	[**+masc**, $-fem$]	[**−masc**, $-fem$]

Moreover, Opitz et al. also found processing differences in a LAN-region between the three genders in German when they compared the processing of well-formed phrases (i.e., a stronger negative amplitude for masculine nouns).[4] There are at least two crucial points with respect to this observation. The first is that there is an observable effect on the processing of well-formed phrases, depending on the gender of a noun. The second point is that Opitz and colleagues provide a possible explanation of the observed effect in terms of a (rather simple) processing model of the checking of grammatical agreement relations. The checking process was modeled as a simple comparison of two feature sets: the set of features of the adjective is compared with that of the noun.[5] The more features are involved, the higher the amplitude of the LAN for

[3]Conflicting features set in bold face.

[4]Note that the phrases for each gender category were matched regarding frequency, length, plausibility, and familiarity.

[5]This, of course, is also the stage in which agreement errors are detected. These errors can in principle be of two types: a) both sets contain conflicting feature specifications, e.g. [+masc] and [−masc], or b) another sub-process (that has to be assumed by Opitz and colleagues but is

the processing of correct phrases. The assumed feature sets for well-formed phrases are given in (2).

(2) Sizes of feature sets in well-formed NPs before A-N agreement (following Opitz et al. 2013: 257)

a. durch schlichten Geschmack
   by plain$_{Masc}$ taste$_{Masc}$
   $[+masc, -fem]$ $[+masc, -fem]$
   comparison of many features → strongest LAN

b. durch schlichtes Design
   by plain$_{Neut}$ design$_{Neut}$
   $[-fem,]$ $[-masc, -fem]$
   comparison of fewer features

c. durch schlichte Struktur
   by plain$_{Fem}$ structure$_{Fem}$
   $[\quad]$ $[-masc, +fem]$
   comparison of fewest features → weakest LAN

However, there is an important objection that could be raised against the reasoning proposed by Opitz et al. The critical word on which the effect was observed was the noun. Although all nouns were matched regarding length and frequency, one cannot exclude the possibility that the observed differences were caused by other factors inherent to these nouns themselves.

3. The present study

The starting point of the present study was primarily the attempt to address the above mentioned objection concerning the results and the reasoning in Opitz et al. (2013), respectively. Thus, we raised the following question underlying our research: Do we find differences in the processing of German nouns between grammatical genders that are not due to a syntactic process of agreement

not spelled out in detail) detects that there is another marker in the system that is more specific and that should have been chosen.

checking? If so, can the assumption of underspecification of grammatical features be maintained?

Three experiments were carried out in order to address this issue. The first experiment examines differences between nouns of different genders when they are embedded in short phrases. In contrast, Experiments 2 and 3 examine the processing of bare nouns without any syntactic context. In addition, the influence of word frequency and morphological structure is addressed, respectively.

3.1. Experiment 1 – Gender and Agreement

The first experiment is basically an attempt to replicate the findings of Opitz et al. (2013) using a behavioral experimental paradigm. Due to the technical requirements of the design no reaction times were registered in the original ERP experiment. Nevertheless, reaction times as well as error rates may reliably indicate processing difficulties. Experiment 1 thus served as an additional control whether behavioral experimental designs such as grammaticality judgments and lexical decisions yield significant data sensitive enough to reveal effects that are supposed to be grounded in very subtle variations of abstract grammatical features. Thus we expected to find analogous results mirroring the different amplitudes of LAN in Opitz et al. (2013) obtained with grammatically well-formed phrases of the three genders.

3.1.1. Method

Participants

A total of 30 German native speakers were tested. Their age ranged from 20 to 39 (mean age of 28 years). Eighteen of them were female. All were payed for their participation.

Materials

The same materials as in Opitz et al. (2013) were used, comprising of 180 German nouns (60 belonging to each of the three genders), each of which was embedded in a syntactic structure of the type Preposition + Adjective + Noun (see examples in (3)).

(3) Examples of critical items in Experiment 1 (well-formed phrases)

a. gegen geplanten Transfer
against intended_{Masc} transfer_{Masc}

b. gegen geplante Zensur
against intended_{Fem} censorship_{Fem}

c. gegen geplantes Logo
against intended_{Neut} logo_{Neut}

All nouns were matched for frequency (frequency class 11 and 12 according to http://wortschatz.uni-leipzig.de), length (an average length of six letters), and syllable structure (only disyllabic words), differing only in gender. For the morphosyntactic violation conditions, for each correct phrase two additional incorrect versions were created by combining them with illicit morphological gender markings on the adjective (e.g. *gegen* $\textit{geplantes}_{Neut}$ $\textit{Transfer}_{Masc}$ and *gegen* $\textit{geplante}_{Fem}$ $\textit{Transfer}_{Masc}$ for the grammatically correct form in (3a)). A total of 540 items (180 correct, 360 incorrect) were distributed over three experimental lists, each list containing 60 correct and 120 ungrammatical phrases. In order to balance the number of correct and incorrect phrases, 60 additional correct filler phrases of the same structure as the experimental phrases were constructed and added to each list. In sum, each list contained 240 phrases. The order of items within each list was pseudorandomized for each participant with the following constraints: no more than three items of the same gender or three items of identical grammaticality status were allowed to follow each other consecutively.

Procedure

Each participant was administered to one experimental list and was tested individually. Experimental sessions were run using the EPrime software suite. At the beginning participants received a written instruction in which they were told that they would have to judge presented phrases with respect to their grammaticality (two choice forced decision). They were instructed to respond as fast and accurate as possible. Each trial started with a fixation sign that was displayed on the screen for 500 ms. After that the phrases were presented visually word-by-word centered on a computer screen with a duration of 300 ms per word. As soon as the noun appeared participants could make their judgment by pressing a corresponding yes or no button.

	feminine	neuter	masculine
RT in ms	738	769	777
% of correct responses	96.6	92.9	90.9

Table 2: Results for grammatical phrases of Experiment 1

3.1.2. *Results*

Only data of well-formed phrases (intended yes-answers) are reported here. Error rates for each gender as well as reaction times of correct responses are given in Table 2 and illustrated in Figure 1.
Accuracy of responses was further analyzed using a one-way ANOVA with the factor Gender (feminine, neuter, masculine). This test showed a significant influence of Gender on the accuracy of responses: $F1(2,12) = 2.95$, $p < .05$; $F2(2,177) = 4.94$, $p < .01$. A statistical post hoc test (Scheffé, critical diff. = 4.8%) revealed that feminine phrases were rated with the highest accuracy (96.9%) and masculine phrases with the lowest accuracy (90.9%). Neuter nouns scored in between and differed statistically neither from masculine nor feminine phrases.

Reaction time data were only analyzed for correct responses to grammatical phrases. First, responses were checked for outliers. All data points that were outside a range of ± 3 standard deviations of a participant's mean were considered as outliers and removed from further analyses. For the remaining data ANOVAs were computed yielding a significant influence of Gender on reaction times: $F1(2,12) = 4.59$, $p < .05$; $F2(2,177) = 3.27$, $p < .05$. A post hoc test (Scheffé, critical diff. = 38 ms) revealed that decisions for feminine phrases were fastest (738 ms) and for masculine phrases slowest (777 ms). Again, neuter nouns scored in between and differed statistically neither from masculine nor feminine phrases.

3.1.3. *Discussion*

Our results provide evidence for processing differences between the three grammatical genders in German. Both the analyses of accuracy rates as well as of latencies in a grammaticality judgment task strongly indicate that the task was more difficult to perform when phrases presented to participants contained

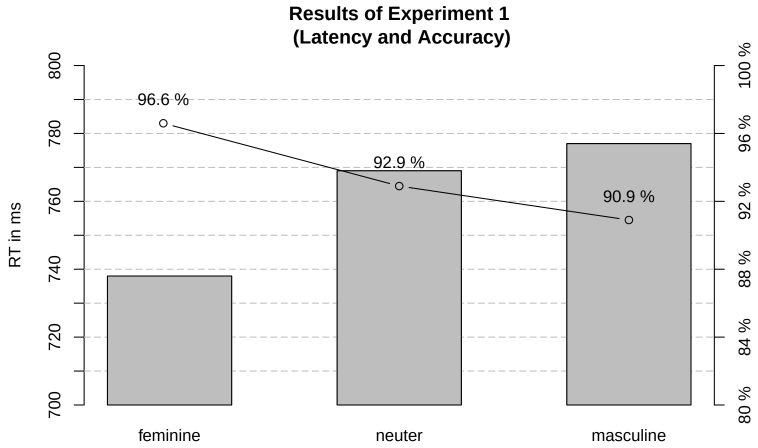


Figure 1: Results for grammatical phrases of Experiment 1

a masculine noun compared to a situation when it contained a feminine noun. Recall that this effect emerged in participants' responses even to perfectly well-formed phrases. As the materials were carefully controlled for length, frequency and plausibility of the phrases, these factors can be ruled out as source of the effect. Therefore, we conclude, in accordance with the explanation provided in Opitz et al. (2013) (see above) that the observed difference was caused by the inherent parsing process that has to check grammatical agreement in order to build up a adequate structure of the incoming phrase.

The second important finding is a methodological one. The results perfectly mirror the findings of Opitz et al. (2013) who obtained a stronger LAN for well-formed phrases with masculine nouns compared to phrases with feminine or neuter nouns. Thus, behavioral methods as the measurement of reaction times and accuracy rates are sensitive enough even for subtle grammatical processes assumed to be the underlying cause of the effect.

However, at least two objections may be raised against our proposed explanation for the observed pattern of responses. In the present study as well as in Opitz et al. (2013) participants had to perform a rather unnatural task. It might be that this kind of meta-linguistic task triggers cognitive processes and strategies not normally involved in language processing, although it would be unclear why masculine forms should be more affected by these processes and strategies then feminine forms. The second objection is that one cannot reliably conclude that it is indeed the matching process (comparing differently large feature sets) that causes the effect and not, instead, some other properties inherent to the nouns themselves.

In order to address these questions, two further experiments were carried out.

3.2. Experiment 2 – Gender and frequency

In the second experiment we used a very common behavioral task that has been extensively used to investigate different aspects of lexical retrieval and lexical representation: lexical decision. Furthermore, the main question in this experiment was whether we could identify differences between the three gender categories of nouns themselves. As an additional factor the frequency of the nouns was systematically manipulated.

3.2.1. Method

Participants

Twenty-three German native speakers were tested. Their age ranged from 19 to 29 (mean of 24 years).

Materials

A total of 144 nouns were used as experimental items, 48 of each of the three genders. In addition, 24 items in each group were of relative low frequency (class 13 according to http://wortschatz.uni-leipzig.de) and 24 were of relative high frequency (class 8). All items were controlled for length. In addition 144 pseudowords were created that served as fillers in order to balance the yes and no responses.

Procedure

Participants were tested individually. The items were presented visually on a computer screen in a pseudo-randomized order. Participants were instructed

	neuter	feminine	masculine	
high freq.	547	557	572	559
low freq.	586	588	591	588
	567	573	582	

Table 3: Results of Experiment 2 (in ms)

to decide as fast and accurately as possible whether the presented word was an existing word of German or not by pressing one of two response buttons.

3.2.2. Results

Responses were checked for outliers. All data points that were outside a range of ± 3 standard deviations of a participant's mean were considered as outliers and removed from further analyses. Statistical tests were performed on the remaining latency data of correct yes-responses using ANOVAs. Main effects were obtained for Gender ($F(2,44)= 4.56$, $p<.05$) and Frequency ($F(1,22)= 59.85$, $p<.001$) as well as a tendency towards an interaction of both factors ($F(2,44)= 2.77$, $p<.07$). This interaction was resolved by two separate analyses for the high frequency and low frequency group, respectively, revealing no effect of Gender for low frequency nouns ($F(2,44)=0.29$, $p=.75$) but a highly significant effect for high frequency nouns ($F(2,44)=12.87$, $p<.001$). Post hoc tests (Scheffé, crit. diff. = 13 ms) for the nouns of high frequency revealed that it took longer to accept masculine (573 ms) nouns than feminine (557 ms) or neuter nouns (547 ms). The results are summarized in Table 3 and illustrated in Figure 2.

3.2.3. Discussion

The first result of Experiment 2 is quite clear: nouns of higher frequency are recognized faster (559 ms) than words of lower frequency (588 ms). This is a trivial finding. Two other findings, in contrast, were quite surprising. Not only were responses influenced by the noun's gender, but this influence was only significant for nouns of relatively high frequency. Again it was the group of masculine nouns that yielded the longest latencies compared to feminine and neuter nouns (which did not differ from each other statistically).

Thus, the results are to be taken as a first indication that processing differences reported in former experiments may not solely be caused by a grammatical checking process. Instead, some properties of the nouns themselves cause a measurable delay in the processing of masculine forms. Moreover, this influence seems to be stronger for nouns of high frequency than for nouns of low frequency. A first explanation of this unexpected frequency and gender interaction could be that, if lexical decision is considered a complex process involving many different sub-processes, the impact of one such sub-process (i.e., retrieval of gender) on the overall task performance is stronger if the impact of other sub-processes (frequency) is reduced. Thus, the higher demands on cognitive resources for nouns of lower frequency obscure the effects of more subtle processes like those underlying the gender effect. For further discussion see section 4.

However, it could be objected that the inherent property of the nouns that seems to cause the effect does not necessarily need to be their gender information. It is at best unclear whether information about the gender of a

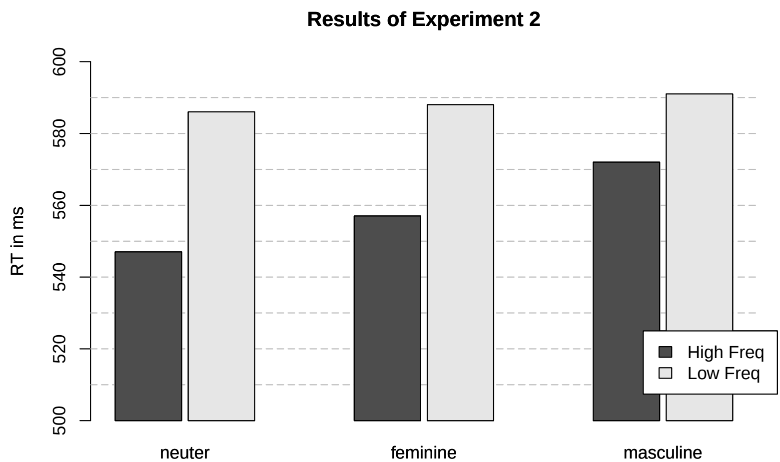


Figure 2: Results for grammatical phrases of Experiment 2 (in ms)

noun is accessed during a lexical decision task at all. To address this issue, a third experiment was carried out.

3.3. Experiment 3 – Gender and morphological marking

Results of Experiment 2 indicated that some inherent properties of a noun (that are at least correlated with their gender) influence their processing in lexical decision. To be sure that information about the gender is accessed during the task, but–crucially–without using any phrase structure yielding agreement, we decided to employ another task, namely gender verification. In this kind of task participants have to decide whether a certain noun is a member of a given gender category or not.

Furthermore, we addressed the question whether the morphological structure of the noun influences this decision. The rationale behind this was the fact that certain derivational suffixes in German clearly determine the noun's gender (e.g., *-ung* for feminine, *-ling* for masculine).[6]

3.3.1. Method

Participants

A total of 18 German native speakers were tested. Their age ranged from 22 to 32 (mean 26 years).

Materials

Only feminine and masculine nouns were used. This was due to the fact that the task should be executable by pressing one of two buttons ('Is the presented noun masculine (Button A) or feminine (Button B)?') and because there are far less derivational suffixes for neuter than for feminine and masculine nouns. A total of 84 nouns were chosen, 42 feminine and 42 masculine. In each of these two groups there were 21 mono-morphemic nouns and 21 nouns with derivational affixes clearly indicating their gender. All nouns were controlled for length (mean length: 7 graphemes) and frequency (mean frequency class 13).

[6]Note that in Experiment 1 as well as 2 roughly 20%-30% of the nouns within the group of each gender comprised of such derivational affixes. Thus, it could not be excluded that these particular morphological cues (or structures) were responsible for the observed effects.

	feminine	masculine	
morph. marked	707	768	738
moprh. unmarked	723	770	747
	715	769	

Table 4: Results of Experiment 3 (in ms)

Procedure

Participants were tested individually. Items were presented visually and in a pseudo-randomized order. Participants' task was to decide whether the presented word was masculine or feminine by pressing a corresponding button. Latency of their decisions was recorded and later analyzed.

3.3.2. Results

Data were scanned for outliers and all data points that deviated more than three standard deviations from a participant's mean were excluded. The remaining data were analyzed using an ANOVA, yielding a main effect for Gender ($F(1,17)=4.98$, $p<.001$) but no effect for Morphological Marking (i.e., presence or absence of a derivational affix) nor any interaction (all p-values >.20). The results are summarized in Table 4 and illustrated in Figure 3. As can be seen, the difference between masculine and feminine was 54 ms with masculine nouns, again, yielding significantly longer reaction times. We found no evidence for any influence of morphological marking.

3.3.3. Discussion

The results of Experiment 3 are in principle in line with the findings of all previous experiments. It was again the group of masculine nouns that yielded longer reaction times. As the particular task of the experiment required gender information of nouns to be accessed, it is now even more likely that the obtained effect in this and the former experiments was indeed caused by the gender features themselves. Interestingly, we did not obtain any evidence indicating that morphological gender marking influences retrieval and/or checking of gender features.

4. General discussion

In all three experiments we found evidence that gender features of nouns have a measurable impact on language processing in German. This influence was found both when the nouns were embedded in grammatical phrases and when they were processed as bare nouns. It was consistently the case that masculine nouns (or phrases containing them) induced longer reaction times and lower accuracy rates, both indicating increased processing demands for theses nouns compared to members of the neuter and feminine category.

The starting point of this study was a verification and replication of findings of an ERP experiment reported in Opitz et al. (2013). There it was argued that the observed effects on the processing of phrases were caused by underspecified inflectional markers and a parsing process that checks grammatical agreement.

In our interpretation of the present data we want to maintain the underspecification hypothesis. In order to do so and to find a unified explanation for the former ERP data as well as the new empirical findings, two new assumptions are necessary that are crucial. The first is the very central point of this paper. We

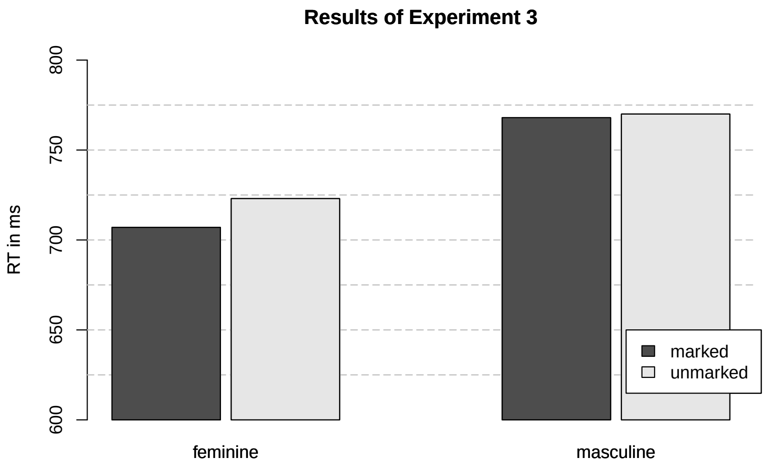


Figure 3: Results for grammatical phrases of Experiment 3 (in ms)

assume that underspecification is not restricted to the domain of inflectional markers but, instead, is a fundamental principle for the lexical representation of nouns themselves. Therefore we propose the following preliminary specification of gender features for German nouns in the mental lexicon, as given in (4). We assume that masculine nouns comprise of the most features, i.e. two features specifications, and that they are thus the most specific forms. In contrast, neuter and feminine nouns comprise of fewer features. At the point being, let us assume that both lack at least one feature. They are therefore underspecified regarding their grammatical gender information.[7]

(4) Lexical specification of gender features in the mental lexicon version A:

a. masculine nouns ⟷ [+masc, −fem]
b. neuter nouns ⟷ [−fem]
c. feminine nouns ⟷ [+fem]

The second new assumption concerns the formerly assumed process of feature matching for the evaluation of grammatical agreement. The precise mechanism of this process, as proposed in Opitz et al. (2013), cannot be maintained without modification.

In the following section we will provide a sketch of this modified mechanism, followed by some considerations regarding the assumed specification of gender features of nouns.

As just mentioned, the evaluation process that checks for grammatical agreement while parsing a phrase has to be modified and worked out in more detail in order to explain the pattern of empirical data. It will be shown later that the evaluation process is not crucially challenged with regard to the processing of grammatically *well-formed* phrases (see below). But, however, in the case of *ungrammatical* phrases it cannot any longer be argued that the observed differences in Opitz et al. (2013) between two different violation conditions are caused by a qualitatively different violation of Compatibility versus Specificity because there is no conflicting feature [−masc] present in the specification of neuter nouns anymore.

However, we suppose that the observed differences in the LAN amplitudes

[7]Note that, at this point, the supposed number of features roughly corresponds with the empirically indicated processing demands. A more detailed explanation for this particular configuration as well as a further modification is given below.

can still be explained while maintaining the assumption of underspecification. In the crucial illicit cases for neuter nouns, a system with underspecified gender features as in (4) would yield configurations like in the examples in (5). (Note that for the sake of clarity case and number features are omitted here, as they do not interfere with the evaluation of gender features).

(5) Two types of illicit neuter agreement and related LAN effects, modified version (following Opitz et al. 2013: 242)

a. *correct*

durch	schlichtes	Design
by	plain_{Neut}	design_{Neut}
	$[-fem]$	$[-fem]$

b. *incorrect 1*
('compatible', excluded by Specificity) →LAN

*durch	schlichte	Design
by	plain_{Fem}	design_{Neut}
	[ ]	$[-fem]$

c. *incorrect 2*
(no longer 'incompatible', thus no violation of Compatibility)[8]
→stronger LAN

*durch	schlichten	Design
by	plain_{Masc}	design_{Neut}
	[**+masc**, $-fem$]	$[-fem]$

Evidently, under such an analysis there is no longer a conflict of feature values between the adjective and the noun in (5c). However, the observed asymmetry between the two illicit phrases in terms of different LAN amplitudes can still be accounted for. But instead of presuming a violation of the Specificity versus the Compatibility criterion, the source of the effect has to be shifted to other sub-parts of the mechanism of the feature checking process.[9] Instead of just comparing two feature sets as a whole and detecting mismatches (like

[8] Formerly conflicting features set in bold face.
[9] And thereby unifying the cause of this particular effect and the cause of the LAN effect for well-formed phrases, see below.

in a unification process), the feature checking could be better modeled in a two-stage or bi-directional way. At first, every feature of a previously parsed word (e.g. [+masc, −fem]) is looking for a corresponding feature on the new incoming word that has to be integrated into the structure (e.g. [−fem]). If a given feature (like [+*masc*]) does not find its counterpart, then this yields a severe violation. Vice versa, every feature of the new word has to search for a corresponding feature in the previously parsed structure. There are in principle at least two ways in order to achieve the observed asymmetry between the two violation conditions. It could either be that a) these two sub-processes operate consecutively and thus the failure of the first search is a more immediate disturbance of the parsing process, or b) both searches may operate simultaneously but a violation of the search from the already parsed structure to the incoming material is regarded as more severe than a violation of the search in the opposite direction for independent reasons. Nevertheless, the observed asymmetry in LAN responses then should be due to the *different directionality* of the two searches. As can be seen in (5c), this particular configuration now yields a major violation because the feature [+masc] does not find a corresponding feature associated with the incoming word *Design*. In contrast, in (5b), there is no gender feature in the structure so far that could search for a corresponding feature in the incoming word. Thus the first search runs without any complications. Contrary, the feature [−fem] associated with the incoming noun *Design* now searches for a corresponding feature in the available structure. As it does not find an appropriate matching feature this, in turn, yields a minor violation (as mirrored in a smaller LAN amplitude). Clearly, this issue has to be addressed in more detail in further research.

But, besides these minor complications, as mentioned above, the empirical results of the processing of well-formed phrases and single words can be uniformly accounted for by this revised feature checking process together with the notion of underspecified gender features of nouns.

If one looks at the matching process for well-formed phrases with a reduced, i.e. underspecified, inventory of gender features for nouns, it becomes readily obvious that the process is still operational and, moreover, is still compatible with the observed LAN effects as well as with the pattern of behavioral results. Even under the revised version of the matching process and the assumption of underspecified gender features of nouns, it is still the case that the number of checking operations (number of searches for features) corresponds with the observed changes in amplitude of LAN as well as increased reaction time and reduced accuracy. This is illustrated in (6). Again, case features are omitted here for the sake of the argument, as they do not interfere with the checking of gender features.

(6) Matching process with underspecified gender features, well-formed phrases

a. durch schlichten Geschmack
by plain_{Masc} structure_{Masc}
$[+masc, -fem]$ $[+masc, -fem]$
comparison of many features: 2 by 2 → strongest LAN
→ longest reaction times
→ lowest accuracy

b. durch schlichtes Design
by plain_{Neut} structure_{Neut}
$[-fem]$ $[\quad -fem]$
comparison of fewer features: 1 by 1

c. durch schlichte Struktur
by plain_{Fem} structure_{Fem}
$[\quad]$ $[\quad (+fem)]$
comparison of fewest features: 0 by 1 → weakest LAN
→ shortest reaction times
→ highest accuracy

Beyond the fact that the number of features still mirrors processing load, there is another striking observation: *none* of the features of the inflectional markers makes use of any of the features that are omitted in the underspecified nouns. This finding does not seem to be accidental. In (7) the whole inventory of markers of the maximally underspecified analysis in Opitz et al. (2013) is listed. None of the markers there realizes any feature [−masc].

(7) Exponents of the strong pronominal/adjectival inflection in German *maximally underspecified account*, (Opitz et al. 2013: 244):

entry of lexical item	*morpho-syntactic context of occurrence*
/n/ ⟷ [+pl,+obj,+obl]	(dat.pl.)
/m/ ⟷ [−fem,+obj,+obl]	(dat.masc.sg./neut.sg.)
/s/ ⟷ [−fem,+obl]	(gen.masc.sg./neut.sg.)
/r/ ⟷ [+obl]	(dat./gen.fem.sg., gen.pl.)
/n/ ⟷ [+masc,−fem,+obj,−obl]	(acc.masc.sg.)
/r/ ⟷ [+masc,−fem]	(nom.masc.sg.)
/s/ ⟷ [−fem]	(nom./acc.neut.sg.)
/e/ ⟷ [ ]	(nom./acc.fem.sg./pl.)

In other words, no morphological process that instantiates an inflected form ever makes use of the feature [−masc]. In a striking parallel, no such feature has to be assumed as part of any lexical representation of nouns, as supported by our empirical findings. In other words, the feature [−masc] can safely be omitted in nouns without inducing failures in the process of inflectional marking.

Thus, while the results in Opitz et al. (2013) support the notion of a maximally underspecified marker inventory instead of an minimally underspecified one, the present findings indicate that this demand for an economically organized lexicon, that stores only as much information as necessary, extends from the inventory of inflectional markers to the lexicon as a whole: underspecification is a *global* principle of the mental lexicon.

Together, the present empirical findings and the results of Opitz et al. (2013) indicate that the lexicon is organized in a highly efficient way. Of course, more research has to be carried out on this issue in order to exclude further possible confounds and to investigate whether the principle of global underspecification extends to other domains or can be found in other languages.

Another aspect that is worth investigating is the concrete role and specification for feminine and neuter nouns. In almost all results so far, masculine nouns showed clear effects of contrast to feminine and/or neuter nouns. The relation between the latter, on the other hand, is less clear as they tended not to differ statistically from each other. For this reason, both are associated with one gender feature each in the preliminary specification in (4), which is in contrast to two such features for masculine nouns.

But, however, if one looks closer at the inventory of markers in (7), it becomes obvious that [−masc] is not the only missing binary feature value there. In addition, the inventory also makes no use of the feature value [+fem]. And, again, it is a most striking finding that the whole checking mechanism would still be completely operational if none of the nouns would carry any such feature. This can be regarded as a first argument that [+fem] is not necessary for gender specifications of nouns. A second argument for this assumption is the case of well-formed feminine phrases. Imagine a case in which [+fem] is present at the nouns like it is illustrated in (6c). The first stage of the process (i.e., the first search) would yield no complications at all, as there are no (gender) features present in the structure already parsed. But, however, after that, a secondary search, which would look for the features [+fem] associated with the noun in the already parsed structure would fail. This, in turn, should lead to a violation and a defective parsing. This situation can be avoided, however, if feminine nouns even lack the feature [+fem]. If there is no such feature on the noun in the first place, it could not trigger a search and therefore no parsing problem (missing of a corresponding feature in the already parsed structure) would occur.

Thus, if even [+fem] is removed from lexical specifications of nouns in total because it is never part of any specification of inflectional markers (and therefore never used for morphological operations), one would end up with the following gender specifications for nouns in (8), making feminine gender the absolute default of the system.

(8) Lexical specification of gender features in the mental lexicon, version B:

a. masculine nouns ⟷ [+masc, −fem]
b. neuter nouns ⟷ [−fem]
c. feminine nouns ⟷ [ ]

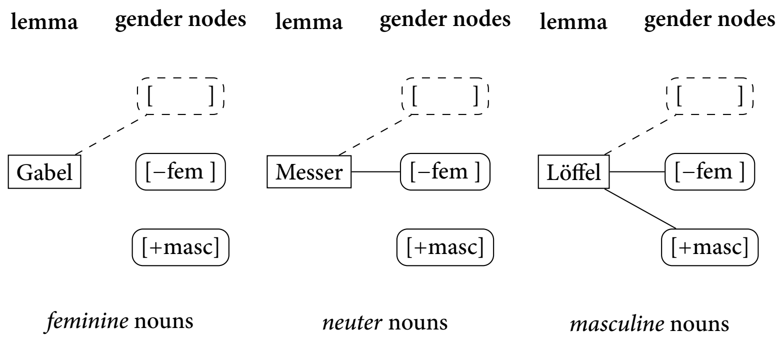
A final remark concerns the question how such a lexical specification as in (8) can cause the observed differences for the processing of bare nouns as in, for instance, lexical decision and gender verification tasks. The preliminary answer would be that lexical retrieval of less specific information, as for feminine nouns, should be less costly than the retrieval of more specific and therefore more complex information. Given the specifications in (8) one could alternatively think of this process as the activation of hierarchically dependent features (nodes). This idea is illustrated in the inheritance tree in (9).

(9) Lexical specifications of gender features in the mental lexicon, modeled as hierarchically dependent nodes:

$[\quad]_{=feminine}$
|
$[-\text{fem}]_{=neuter}$
|
$[+\text{masc}]_{=masculine}$

Or, alternatively, gender representations can be modeled as generic gender nodes in an activation based model (see, e.g., Levelt et al. 1999). In such a case, nouns of different gender categories would differ in their association to these generic gender nodes. This idea is illustrated in (10).

(10) Lexical specification of gender features in the mental lexicon, represented as generic nodes
examples for feminine (*Gabel*), neuter (*Messer*), and masculine (*Löffel*) nouns



The retrieval of feminine nouns would not necessarily include activation of any gender features, as indicated by the gender node comprising of an empty feature set. Moreover, it is not clear whether the activation (or even existence) of such an empty node would be necessary at all, as indicated by the dashed lines. On the other hand, retrieval of neuter nouns would demand the activation of at least a [−fem] node. Finally, the retrieval of masculine nouns would require both, the activation of [−fem] and [+masc]. Thus, the least amount of activation is needed for the retrieval of feminine nouns, more activation for the retrieval of neuter nouns, and, finally, the retrieval of masculine nouns would demand most activation. This architecture would have implications, for instance, for priming experiments.

Of course, all these assumptions and implications have to be further tested empirically in order to support the hypothesis of global underspecification. Beyond additional empirical evidence, our assumptions need to be supported by more grounded theoretical considerations and implementations into parsing models. At least at the point being, we could think of one supporting theoretical argument in favor of a global underspecification hypothesis. There is good reason to assume that if a principle–like underspecification–is an essential part of the architecture of a cognitive system (for which there is a wealth of evidence in theoretical morphology in the domain of inflectional), this principle might be used more universally in a large range of domains. Note also that there is empirical evidence for conceptually related notions of underspecification in other domains of the grammar as well (see, for instance, Frisson and Pickering (1999), Pickering and Frisson (2001) for semantic underspecification, or the theory of the Featurally Underspecified Lexicon (FUL) in phonology (cf. Lahiri and Reetz 2010, 2002)).

5. Conclusion

In this paper we have presented empirical evidence that lexical representations of German nouns may be underspecified with regard to abstract gender features. In a series of experiments it was consistently found that masculine nouns show indications of an increased processing load compared to feminine or neuter nouns. We assume that the observed effects are due to an underspecified representation of grammatical features. We propose that underspecification is more broadly used by the human language system than normally assumed and that it might be a *global* characteristic of the mental lexicon. The mental lexicon is organized in a highly efficient and economic way. Correspondingly, if a grammatical feature is never used for morphosyntactic operations (as, e.g., [−masc] in inflectional marking in German) it is probably not represented in the system at all. Thus, the human language system may be considered as optimal from an optimal design's perspective. It reduces both the demands for storage (by reducing the amount of stored information) and for processing component (by not taking into account redundant information in online evaluation processes and therefore keeping the number of required sub-processes, i.e. search operations, to a minimum) (cf. Chomsky 2005).

We further introduced a modified version of a basic parsing model for the evaluation of grammatical agreement that was originally proposed in Opitz et al. (2013). In order to make this parsing model compatible with both the neurophysiological evidence reported in Opitz et al. (2013) and the behavioral results of the present paper, we assume it to comprise of a two-directional search for corresponding features between new, incoming elements and already parsed structures.

However, it is clear that all assumptions and conclusions provided in this paper are of preliminary nature. Further empirical evidence is needed to support our conclusions and, moreover, the results will have to be implemented into theoretical models of inflection and language processing in more detail.

References

Bierwisch, Manfred (1967): Syntactic Features in Morphology: General Problems of So-Called Pronominal Inflection in German. In: *To Honour Roman Jakobson*. Mouton, The Hague/Paris, pp. 239–270.

Blevins, James (1995): 'Syncretism and Paradigmatic Opposition', *Linguistics and Philosophy* **18**, 113–152.

Bybee, Joan (1985): *Morphology: A study of the relation between meaning and form*. John Benjamins, Philadelphia, PA.

Bybee, Joan (1995): 'Regular Morphology and the Lexicon', *Language and Cognitive Processes* **10**(5), 425–455.

Caramazza, Alfonso, Alessandro Laudanna and Cristina Romani (1988): 'Lexical access and inflectional morphology', *Cognition* **28**, 297–332.

Caramazza, Alfonso, Gabriele Miceli, M. Caterina Silveri and Alessandro Laudanna (1985): 'Reading mechanisms and the organization of the lexicon: Evidence from acquired dyslexia', *Cognitive Neuropsychology* **2**(1), 81–114.

Chomsky, Noam (2005): 'Three Factors in Language Design', *Linguistic Inquiry* **36**(1), 1–22.

Clahsen, Harald, Sonja Eisenbeiss, Meike Hadler and Ingrid Sonnenstuhl (2001): 'The mental representation of inflected words: An experimental study of adjectives and verbs in German', *Language* **77**, 510–543.

Corbett, Greville and Norman Fraser (1993): 'Network Morphology: A DATR Account of Russian Nominal Inflection', *Journal of Linguistics* **29**, 113–142.

Frisson, Steven and Martin J. Pickering (1999): 'The processing of Metonymy: Evidence from eye movements', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **25**, 1366–1383.

Gallmann, Peter (2004): Feature Sharing in DPs. *In:* G. Müller, L. Gunkel and G. Zifonun, eds, *Explorations in Nominal Inflection*. Mouton de Gruyter, Berlin, pp. 121–160.

Grimshaw, Jane (2001): Optimal Clitic Positions and the Lexicon in Romance Clitic Systems. *In:* G. Legendre, J. Grimshaw and S. Vikner, eds, *Optimal Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 205–240.

Halle, Morris and Alec Marantz (1993): Distributed Morphology and the Pieces of Inflection. *In:* K. Hale and S. J. Keyser, eds, *The View from Building 20. Essays in Linguistics in Honor of Sylvain Bromberger*. Vol. 24 of *Current Studies in Linguistics*, MIT Press, Cambridge, Mass., pp. 111–176.

Halle, Morris and Alec Marantz (1994): Some Key Features of Distributed Morphology. *In:* H. H. Andrew Carnie and T. Bures, eds, *MITWPL: Papers on Phonology and Morphology*. Vol. 21, MITWPL, Cambridge, pp. 275–288.

Lahiri, Aditi and Henn Reetz (2002): 'Underspecified recognition', *Labphon* 7, 637–676.

Lahiri, Aditi and Henning Reetz (2010): 'Distinctive Features: Phonological underspecification in representation and processing', *Journal of Phonetics* **38**, 44–59.

Levelt, Willem J. M., Ardi Roelofs and Antje S. Meyer (1999): 'A theory of lexical access in speech production.', *Behavioral and Brain Sciences* **22**(1), 1–75.

Lukatela, Georgije, Branislaw Gligorijević, Aleksandar Kostić and Michael T. Turvey (1980): 'Representation of inflected nouns in the internal lexicon', *Memory and Cognition* **8**, 415–423.

Lukatela, Georgije, Z. Mandić, Branislav Gligorijević, Aleksandar Kostić, M. Savić and Michael T. Turvey (1978): 'Lexical decision for inflected nouns', *Language and Speech* **21**(2), 166–173.

Opitz, Andreas, Stefanie Regel, Gereon Müller and Angela D. Friederici (2013): 'Neurophysiological evidence for morphological underspecification in German strong adjective inflection.', *Language* **89**(2), 231–264.

Penke, Martina (2006): *Flexion im mentalen Lexikon*. Max Niemeyer: Tübingen.

Penke, Martina, Ulrike Janssen and Sonja Eisenbeiss (2004): 'Psycholinguistic evidence for the underspecification of morphosyntactic features', *Brain and Language* **90**, 423–433.

Pickering, Martin J. and Steven Frisson (2001): 'Processing ambiguous verbs: Evidence from eye movements', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **27**, 556–573.

Rumelhart, David E. and James L. McClelland (1982): 'An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model', *Psychological Review* **89**, 60–94.

Sauerland, Uli (1996): The Late Insertion of Germanic Inflection. Generals paper, MIT.

Seidenberg, Marc S. and Laura M. Gonnerman (2000): 'Explaining derivational morpholoy as the convergence of codes', *Trends in Cognitive Sciences* 4(9), 353–361.

Stockall, Linnaea and Alec Marantz (2006): 'A single route, full decomposition model of morphological complexity: MEG evidence', *The Mental Lexicon* 1:1, 85–123.

Stump, Gregory (2001): *Inflectional Morphology*. Cambridge University Press, Cambridge.

Trommer, Jochen (2006): 'Person and Number Agreement in Dumi', *Linguistics* 44, 1011–1057.

Wunderlich, Dieter (1996): Minimalist Morphology: The Role of Paradigms. *In:* G. Booij and J. van Marle, eds, *Yearbook of Morphology 1995*. Vol. 99, Kluwer, Dordrecht, pp. 93–114.

Wunderlich, Dieter (1997): Der unterspezifizierte Artikel. *In:* C. Dürscheid, K. H. Ramers and M. Schwarz, eds, *Sprache im Fokus*. Niemeyer, Tübingen, pp. 47–55.

Wunderlich, Dieter and Ray Fabri (1995): 'Minimalist Morphology: An approach to inflection', *Zeitschrift für Sprachwissenschaft* 14, 236–294.