# German Noun Class as a Nominal Protection Device

Richard Futrell

Undergraduate Honors Thesis

Stanford University, Department of Linguistics

May 2010

_____

Faculty advisor: Dan Jurafsky

_____

Second reader: Michael Ramscar

*Gretchen: Wilhelm, where is the turnip?*
*Wilhelm: She has gone to the kitchen.*
*Gretchen: Where is the accomplished and beautiful English maiden?*
*Wilhelm: It has gone to the opera.*

- Mark Twain, "The Awful German Language"

## 1. Introduction

Grammatical gender, also known as noun class, afflicts about half of the world's languages. Speakers of these languages must mark each noun for its membership in a certain noun class, and must similarly mark elements such as adjectives or verbs that agree with the noun. In over half of these cases, the choice of gender for a noun has no comprehensive systematic relationship to the meaning of the nouns (Corbett 2008), posing a significant obstacle for L2 learners (Harley 1979, Tucker et al. 1968: 312). Some have taken grammatical gender to be superfluous: for instance, Maratsos (1979: 235) calls the existence of such a system "excellent testimony to the occasional nonsensibleness of the species." A number of researchers have nonetheless defended noun class against the accusation of uselessness, and I will take up that cause in this paper.

I will examine the gender system of Standard German with an eye toward detecting function. I claim that noun class serves as a sort of 'Nominal Protection Device,' alleviating the linguistic difficulties inherent in nouns by reducing uncertainty about nouns. Noun class markers help language users predict nouns in a number of ways: they predict the form of the noun, they predict the semantics of the noun, and they predict

which discourse referent a pronoun points to in reference tracking (in the sense of Barlow 1992). I will examine these three predictive functions of noun class with three studies, asking two kinds of questions: (1) Are there any ways in which the distribution of genders in German is tailored to these functions? And (2) does the presence of gender affect German speakers' use of nouns? The ultimate question I would like to begin investigating is to what extent the form of gender systems can be explained by the function of gender systems in reducing uncertainty about nouns.

Noun classes predict the form of the noun simply by narrowing the number of nouns that can occur in a specific context. In doing so, they reduce the nouns' entropy, an information-theoretic measure of their uncertainty. In section 2, I examine first the logical space of noun class systems, and show that some possible noun class systems would not always reduce nominal entropy in this manner; but in a corpus study I find that German gender does always reduce nominal entropy.

Another way that noun classes reduce uncertainty about nouns is by cueing for their semantics. For instance, alcoholic drinks in German are nearly all masculine (Köpcke 1982); in certain contexts, the gender marking could be a cue that the following noun will stand for an alcoholic beverage. This function would motivate the existence of semantic clusters in gender assignment. While these clusters have been observed, no study has yet looked at them robustly using a corpus and a complete lexicon. I attempt to provide such a study in section 4, showing that nouns with similar semantics are more likely than chance to be in the same gender class.

A third way in which gender can reduce uncertainty about nouns is by providing reference tracking in pronouns. When we use a pronoun there is always some uncertainty

about what noun, or what discourse referent, that pronoun 'stands for', but mandatory gender marking reduces that uncertainty. Using corpora of spoken language, I present evidence in section 5 that noun classes are distributed in the German lexicon so that frequently co-occurring nouns, which are likely to cause ambiguity in reference tracking, are more likely than chance to be in different genders, and I examine how that configuration arises in the lexicon.

Overall, I argue that noun classes provide an ever-present system of predictive cues for nouns that aids language users in utilizing those nouns. This systematic reduction in nominal uncertainty allows German speakers to make use of rarer nouns more frequently, as I show using three corpus studies in section 3. The results all in all argue for the utility of an information-theoretic account of the function and form of noun class.

### 1.1 What is noun class?

In this section I will survey the typical characteristics of noun class systems and the particular characteristics of the German system, outlining the features that a functional explanation should account for.

In a language with noun class (or gender—I will use the terms interchangeably), elements other than a noun itself that bear some syntactic relation to that noun must systematically change in form, according to the noun. In the words of Hockett (1985: 231): "Genders are classes of nouns reflected in the behavior of associated words." This is more or less the definition adopted by Corbett (1991: 1) and Aikhenvald (2000: 28), *unter andern*. But in a probabilistic world, this definition will not get us far. Nowadays, researchers routinely use the behavior of surrounding words to induce English semantic

classes (see for instance Jurafsky & Martin 2009: 659-667). Yet we do not want to say that English displays a noun class system. Nor can we pin down noun class as 'classes of nouns reflected in the *categorical* behavior of associated words', because agreement is not categorical. When agreement based on syntactic form comes in conflict with agreement based on the momentary semantic properties of the object at hand, the semantically-motivated agreement form often wins (Corbett 1979).

For the purpose of this investigation I will frame noun class systems from a probabilistic perspective. Informally, I consider a language to have a noun class system if it contains some linguistic element that (1) mandatorily co-occurs with nouns in some context, (2) has a relatively small set of forms, but more than one form, and (3) has forms that are almost entirely *predicted* by a subset of the properties of the noun[1]. That is, for all nouns $n_i$ and all particular gender-agreeing elements $G$ in a language, $P(G|N=n_i) \approx 1$.

In terms of information theory, we can say a language has a gender system iff it has a set of nouns $N$ and a set of morphological elements $G$ such that (1) there is a syntactic environment in which $G$ must co-occur with $N$, (2) $H(G|N) \approx 0$, and (3) $0 < H(G) < H(N)$.

The definition is full of hedges—"near zero," "almost entirely predicted," etc.—because a probabilistic definition of noun class acknowledges that a system of noun class agreement is only the extreme end of a spectrum of co-occurrence patterns. For instance, we can probably induce a class of beverages in English nouns by observing the behavior of associated verbs: they tend to become *drink* rather than some alternatives such as *eat.*

[1] I do not want to claim that speakers calculate gender on-line on the basis of semantic or phonological features of nouns (as Köpcke (1981) maintains). Speakers could predict the gender of the noun on the basis of simple memorization/association for each noun.

But $H(V|N = water)$ is nowhere near zero. As $H(X|N)$ approaches zero, the element *X* becomes more and more like a gender-agreeing element.

In practice, noun class tends to assign nouns into 2 to 20 classes[2]. Every noun class system has some semantic base which predicts gender assignment for a frequent subset of nouns. The semantic bases of classes in the world include gender, shape, texture, function, religious status, animacy, and several others (Corbett 1991), but they tend to be based on properties that are stable for certain referents. Shape and animacy are also especially salient features that children tend to overgeneralize in their early language use (Aikhenvald 2000). There are such things as verb classes and adjective classes, but they are comparatively rare in the world: usually it is only nouns that enter into a classification system (Aikhenvald 2000).

### 1.1.1 Noun class in German

The German gender system is manifested in agreement in articles and adjectives, which precede nouns within NPs. German divides all nouns into noun classes traditionally labeled MASCULINE, FEMININE, and NEUTER. Articles and adjectives also mark plurality and four cases, NOMINATIVE (usually for subjects of verbs), ACCUSATIVE (usually for objects of verbs), DATIVE (marking an indirect object or the complement of a preposition), and GENITIVE (marking possession). In this paper, for the most part, I examine noun class only as manifested on definite articles immediately preceding nouns.

Table 1 demonstrates the syncretisms in the marking of the definite article:

[2] Noun *classifier* systems, such as the systems of Mandarin and Japanese, do not exhibit agreement and can include hundreds of classes. They are usually treated separately from gender systems and the choice of a non-generic form is often optional.

| Case | Masculine | Feminine | Neuter | Plural |
|---|---|---|---|---|
| Nominative | **der** | **die** | **das** | **die** |
| Accusative | **den** | **die** | **das** | **die** |
| Dative | **dem** | **der** | **dem** | **den** |
| Genitive | **des** | **der** | **des** | **der** |

**Table 1**. Gender, case, and number marking in German definite articles.

### 1.2 Possible functions of noun class

A number of possible functions for noun classes have been advanced in the literature. I will focus on one promising previous suggestion and some of my own, and conclude by discussing some of the commonalities of these proposed functions.

Nouns are a problem for human language users. They are the most numerous part of speech and arguably carry the most information. In conversation, speakers must be able to identify and differentiate several discourse referents instantiated by nouns. Aside from these discourse difficulties, nouns pose a problem for processing. One of the most common sites for disfluencies in English is at the determiner preceding a noun, and the more complex the noun the more disfluencies occur (Clark & Wasow 1998). Nouns are the most common sites for incorrect lexical retrieval and a host of other speech errors (Vigliocco 1997). Noun class marking may play a part in alleviating these burdens.

#### 1.2.1 Nominal entropy reduction

One reason nouns are difficult is that they are the most diverse part of speech. Processing ease and difficulty have been shown to correlate with the information-theoretic measure of entropy, a measure of uncertainty about what linguistic element will follow in a given context (McDonald & Shillcock 2003). Since there are so many different nouns, the entropy of an English noun in the left-bounded context of a determiner, i.e. the spot marked as ***!*** in (1), will be huge.

(1) Yesterday I visited the ***!*** doctor.

Specifically, the entropy is the weighted average of the information of each noun. Let $n_1,...,n_k$ represent all the possible nouns in English in this context. The entropy of the nouns $N$ is

$$H(N) = -\sum_{i=1}^{k} P(n_i) \log_2 P(n_i)$$

(Shannon 1948). Speakers will have trouble producing the noun, and hearers will have (relative) trouble processing the noun, at least more so than for the other transitions in the sentence.

The cure for high entropy is prediction. The sentence in (2) will be easier for speakers to produce, and easier for hearers to process, if the form of the noun is predicted in context, as in (2):

(2) Yesterday when I went to the hospital I visited the *doctor*.

This is where grammatical gender comes in. Genders, when manifested as agreeing elements before the noun, can be a systematic implementation of the same principle. Consider the German equivalent of (1) in (3):

(3) *Gestern besuchte ich den **!** Arzt*
yesterday visited I the.MAS ***!*** doctor

While the amount of context is the same as in (1), the uncertainty about the following noun is greatly reduced in comparison to (1). The following noun must belong to the MASCULINE noun class. Essentially, the German genders conspire to reduce the entropy of German nouns, at the cost of increased entropy of the article. Work such as Jaeger (2009 ms)'s study of complementizer omission reveals that languages strive to approach a *constant entropy rate*, avoiding peaks and troughs in the entropy of their utterances. Gender makes each noun less surprising and smoothes a potential spike in information, and we are closer to the constant entropy rate that language users cherish (see also Genzel & Charniak 2002).

The same entropy reduction is in play when gender-agreeing pronouns help speakers differentiate discourse referents. A gendered pronoun reduces the entropy of the possible referents and narrows the search space.

### 1.2.2 Facilitating noun recall

Arnon & Ramscar (under review) find in an artificial language experiment that articles bearing grammatical gender information, when learned in context, aid in the recognition, production, and appropriate use of nouns. Subjects correctly produced the artificial nouns correctly more often, and applied them to the appropriate pictures more consistently, when they had learned the nouns from sentences containing articles. These subjects outperformed subjects who learned nouns without gender context. An interesting interpretation of Arnon & Ramscar's result is that gender, by forcing speakers to provide a richer set of cues to a given noun, causes hearers to learn that noun more effectively for later recall and use.

Another study hinting at a possible effect of gender on noun recall is Vigliocco's (1997) study of tip-of-the-tongue events. In cases where they ended up producing the correct noun, Italian speakers had access to the noun's grammatical gender before they could produce is phonological form. In a speaker's production, she may remember the gender of a noun but not the noun itself, and from there reconstruct the specific noun successfully. The same principle can apply in a hearer's memory of an utterance, such that gender facilitates accurate recall of noun use for that hearer.

### 1.2.3 Reference tracking via agreeing pronouns

The most concretely-formulated proposal for a function of noun class is that the assignment of nouns into classes helps language users in reference tracking (Zubin and Köpcke 1986: 173-5, Koval' 1979: 97-8, Heath 1975: 95-6, Eisenberg 2006). As an example, consider the example (4), from Zubin & Köpcke (1986: 173):

(4) *der Krug fiel in die Schale, aber er zerbrach nicht*
the.MAS jug fell into the bowl.FEM but it.MAS broke not
'The jug fell into the bowl, but it (the jug) didn't break.'

The referent of the pronoun *it* is unambiguous, because it must have a MASCULINE referent, which in this case must be the jar, not the bowl.

A noun class serving this function would have many of the properties we do observe in German noun class, and in noun-classes across the world. First of all, since it is nouns that convey reference, we would expect a gender system to latch onto nouns, rather than verbs or adjectives. In addition, we would have a convenient explanation for Greenberg's (1969) Universal #43: "If a language has gender categories in the noun, it has gender categories in the pronoun." Also, optimally, nouns would be distributed so

that referents that tend to be found in the same cobntext receive different gender. We would therefore not want a systematically semantic system for gender assignment; a course-grained, irregular system would be less likely to group together nouns that co-occur. Sex provides a good candidate for such a system: humans appear in many of the same contexts regardless of sex.

The limitation of the approach based on reference tracking is that it only works for gender as manifested in agreeing pronouns (Claudi 1985: 43). Agreement on adjectives would be useful for reference tracking in languages with relatively free word order only; in a language such as German, where an attributive adjective's relation to a noun is determined nearly entirely by word order, the adjective endings cannot serve a disambiguating role in reference tracking. The presence of gender agreement in adjectives might then serve as a continual reminder and cue to hearers and learners about each noun's gender (furthermore, note that German adjectives tend to provide gender information only when that information has not been provided by a definite article). The article seems similarly redundant.

Another problem is that reference tracking doesn't seem to cause enough miscommunication to necessitate the introduction of a system as comprehensive as gender agreement. On the other hand, the presence of disambiguating gender agreement may simply allow Germans to use less outside context to disambiguate pronouns, to use pronouns rather than nouns more often, and to increase the distance between pronouns and the nouns they refer to.

### 1.2.4 Other proposed functions of noun class

One interesting approach to noun class has focused on the subtle distinctions it can provide in discourse, especially when class assignment is not entirely rigid. This is often more applicable to classifier systems than noun class systems, but Robinson (to appear) discusses how gender agreement in German has been used to express subtle social perceptions and trends when the appropriate agreement form is unclear. However, in this paper I will focus on cases where gender assignment is invariant.

### 1.2.5 Commonalities of the proposed noun class functions

What the proposed functions for noun class have in common is that they serve to offset various processing difficulties inherent to nouns: they may aid in tracking reference, in reducing entropy, and in recall. For these reasons, I propose that the system of noun classes generally functions as a Nominal Protection Device (NPD), ensuring the effective use of nouns for language users. The basic *modus operandi* of the NPD is to predict nouns. This could be equally useful for hearers, who abhor linguistic uncertainty, and for speakers, for whom the prediction of nouns in context constitutes nominal recall.

To aid in the reduction of nominal entropy and in reference tracking, noun classes should be distributed in such a way that more general context does not predict the gender of a noun. This is because greater context can thwart the protective aspects of noun class agreement. Consider a language where every location is in the FEMININE gender. Say we hear the fragment (5):

(5) Yesterday I went to the.FEM ...

The gender on the article has not reduced the uncertainty about the following noun, nor provided a cue to the following noun, because we are already fairly certain that

the noun will represent a location. Similarly, in a conversation involving several different locations, referring to a certain location with a feminine-marked pronoun will not reduce uncertainty about what discourse referent is in play. For the Nominal Protection Device to function optimally, in fact, noun class should be assigned in such a way that it is *not* predictable in context. Indeed, as one example, Koval' (1979: 97-8) finds that domestic animals in Fula fall into many different genders.

However, Zubin and Köpcke (1986) demonstrate the existence of large swaths of semantic regularity in the German gender system, and argue that semantic regularity is the norm for gender assignment rather than the exception. Furthermore, Lakoff's 'Domain of Experience Principle' (Lakoff 1986: 15), based on examples from Dyirbal, holds that nouns referring to objects that occur in the same 'domain of experience' will have the same noun class. For instance, fishing hooks and fish might be in the same noun class. Such an assignment system may seem to be counterproductive for the NPD if it functions as I outlined here, because the gender of most nouns would be predicted by the semantic frame in which a speaker is speaking.

However, the existence of semantic regularities need not result in gender information being predicted prematurely in context. Semantic regularities do exist even in the most erratic of gender assignment systems. But as long as noun systems do not cluster in categories that are likely to frame whole swaths of conversations, the genders will not be predicted away, and the semantic associations of gender will provide a semantic cue to the noun they are attached to. Noun class systems are torn between avoiding semantic regularity in some fields and craving it in others; the interplay between clustering and dispersal in noun class systems has yet to be studied systematically.

In this paper I will study whether German noun classes are useful in predicting nouns and whether noun class is distributed through the German lexicon to maximize the usefulness of the NPD. First, I will examine whether German noun classes do reduce nominal entropy, and if that reduction in entropy allows German speakers to use rare nouns more easily. Second, I will examine the semantic distribution of nouns, to determine to what extent noun class reflects semantic regularities. Third, I will examine if noun classes are distributed so that conversational topics do not cause clusters of genders, and how that distribution is achieved.

## 2. Noun class and nominal entropy reduction

The main way that noun class could protect nouns would be by predicting them. In doing so, noun class can reduce the entropy of nouns. While it may seem intuitive that noun classes predict nouns to some extent, there are possible cases in which noun class actually *increases* nominal entropy. In this section, I will address this issue on three points. First, I will prove that noun class, as defined in section 1.1, must reduce *conditional* entropy. Second, I will prove that certain agreement forms in a noun class system can nevertheless increase uncertainty about nouns. That noun class reduces uncertainty about all nouns is thus an empirical question, so as the third part of this section, I will show that German noun class does in fact reduce nominal entropy in all cases.

### 2.1 Noun class implies a reduction in conditional entropy

Here I present a proof that noun classes, as defined information-theoretically in section 1.1, must reduce the conditional entropy of nouns. That is, conditional entropy is a weighted average of entropies based on conditional probabilities. Specifically, the conditional entropy of a variable *X* given another variable *Y* is:

$$H(X|Y) = -\sum_{y} P(y) \sum_{x} P(x|y) \log P(x|y) .$$

The definition of a noun class is reproduced in (6) below.

(6) A language has a noun class system iff, having a set of nouns *N*, there is a set of elements *G* such that:

(a) *G* must co-occur with *N* in some syntactic context,

(b) $H(G|N) \approx 0$, and

(c) $0 < H(G) < H(N)$.

In this section I will consider the idealized case where $H(G|N) = 0$ (that is, where agreement is entirely determined by the form of the noun). The goal is to show that $H(N) > H(N|G)$.

**Proof:** The figure $H(X) - H(X|Y)$ is one way of expressing the *mutual information* $I(X;Y)$, a measure of dependence between two variables (Coombs et al. 1970: 319-320). Being symmetrical, mutual information is also expressible as $H(Y) - H(Y|X)$. Thus $H(X) - H(X|Y) = H(Y) - H(Y|X)$. From the idealized case of (6b), we know that $H(G|N) = 0$, and from (6c) we know that $H(G) > 0$, so $H(G) - H(G|N) = H(G) > 0$. So the mutual information $I(G;N) > 0$, and $H(N) - H(N|G) > 0$. Thus, $H(N) > H(N|G)$.

For the proof to hold, it is not even necessary that $H(G|N) = 0$ precisely. It will work as long as $H(G|N) < H(G)$, that is, as long as nouns tell us anything about agreement forms.

### 2.2 Noun class does not always reduce entropy

A noun class system always reduces conditional entropy, but as far as noun class is concerned, conditional entropy is a weighted average over all agreement forms. It thus remains conceivable that there may be an agreement form $g_i$ in $G$ such that $H(N|G = g_i) > H(N)$: a case where a certain agreement form increases uncertainty about the following noun. In this section I will prove that such a configuration of noun class in a lexicon is possible.

**Proof:** Consider the noun class system of the toy language in Figure 1. In the MASCULINE class, there are five nouns each with a probability $P(N_{MAS}) = 5/50$. In the FEMININE class, there are 25 nouns each with a probability $P(N_{FEM}) = 1/50$.

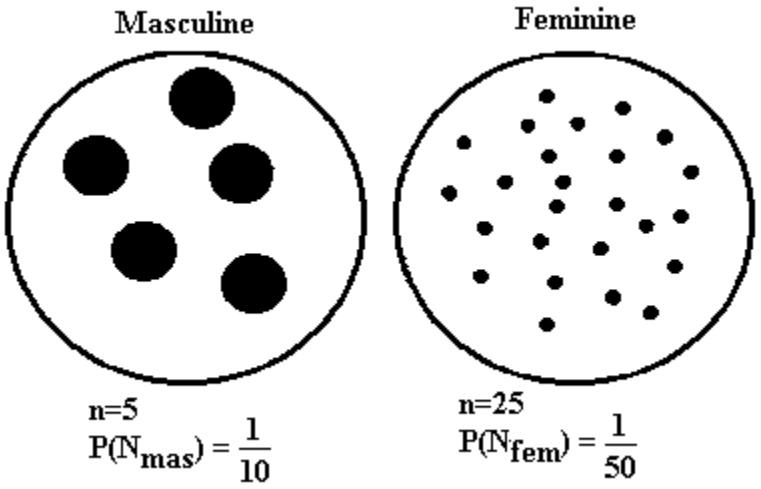


**Figure 1.** Noun classes in a toy language.

At the point when a hearer knows the next word is a noun, but does not know its noun class, his uncertainty is the entropy of all the nouns:

$$H(N) = -\sum_i P(n_i) \log_2 P(n_i)$$

$$H(N) = -(5)\frac{1}{10}\log_2\frac{1}{10} - (25)\frac{1}{50}\log_2\frac{1}{50}$$

$$H(N) = 4.48\ bits.$$

When the hearer discovers that the coming noun will be in the FEMININE class, his uncertainty is the entropy of all the nouns in that class:

$$H(N|G = Fem) = -(25)\frac{1}{25}\log_2\frac{1}{25}$$

$$H(N|G = Fem) = 4.64\ bits.$$

Thus, $H(N|G=Fem) > H(N)$.

The feminine class makes the following nouns *less predictable* by limiting the possible nouns to a large set of low-frequency nouns. This situation arises when one class contains a relatively large number of infrequent nouns.

### 2.3 German noun class reduces nominal entropy

If noun class in German is a device to keep nominal entropy under control, then nouns should be distributed across genders such that this situation does not arise. Given word frequencies from a corpus, it is possible to compute directly whether noun classes bring about a reduction in nominal entropy.

**Method:** The NEGRA II corpus of German newspapers (Skut et al. 1997) was lemmatized using Morphy (Lezius et al. 1998). Frequency data and gender class were collected for each lemma. All contractions involving articles (i.e., *im* for *in dem*) were expanded. For each case, every noun immediately preceded by a definite article was extracted and counted. Using the counts, I calculated the entropy of all the nouns in each case separately:

$$H(N) = -\sum_i P(N_i) log_2 P(N_i)$$
$$= -\sum_i \frac{Count(N_i)}{Total(N)} log_2 P \frac{Count(N_i)}{Total(N)}$$

I then calculated the conditional entropy given each type of article. For a noun following the masculine nominative article *der*, for instance, where $N_{\text{DER}}$ is the set of nouns that can follow *der*, the conditional entropy is:

$$H(N|der) = -\sum_i P(N_i|der) log_2 P(N_i|der)$$
$$= -\sum_i \frac{Count(N_i|der)}{Total(N|der)} log_2 P \frac{Count(N_i|der)}{Total(N|der)}$$

**Result:** As expected (see Figure 2), for no case and no article was the conditional entropy greater than the entropy of all nouns in that case.

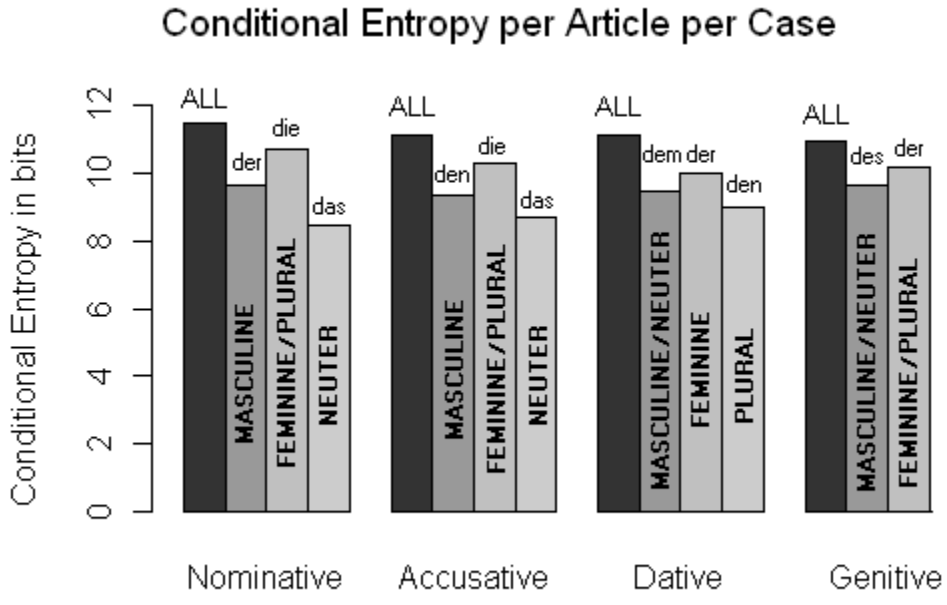


**Figure 2**. Conditional entropy of nouns after articles, per case, and total entropy in each case.

**Discussion:** As long as there is no gender reserved for a myriad of infrequent nouns, then gender reduces entropy. If nouns were assigned to gender in such a way, we would expect noun class to be serving a very different purpose. For instance, if noun class increased the entropy of nouns, it could be used to draw attention to nouns, since surprising high entropy has been shown to attract attention in the visual sphere (Itti & Baldi, 2005).

### 2.4 Summary and Conclusions

The proof in section 2.1 shows that noun class systems must *on average* reduce uncertainty about nouns. However, certain agreement forms within the noun class system may individually increase uncertainty about nouns (section 2.2). This can occur when a noun class belongs solely to a large number of infrequent nouns. While this situation does not arise in German (section 2.3), we do find that individual genders reduce uncertainty about nouns to different degrees.

It is not clear whether the specific entropy reduction brought about by an agreement marker is relevant in language processing, or whether it is the average entropy reduction caused by the system as a whole. Cohen Priva (2008) presents evidence that the entropy of a phoneme in a specific context does not predict its deletion rates; rather, the phoneme's average entropy in all contexts is the best predictor. Similarly, it could be that whatever effect noun class has on nominal behavior is dependent on the average reduction in entropy brought about by the system, rather than the specific entropy reduction caused by a specific agreement marker. By showing that German noun class markers reduce entropy both specifically and on average, I allow myself to remain

agnostic on this question, as both interpretations lead to the same conclusion: German gender reduces uncertainty about nouns.

### 3. Effects of Nominal Protection

In reducing the entropy of nouns in context, a grammatical gender system makes nouns easier to produce and comprehend. This result is shown directly in an artificial language experiment by Arnon and Ramscar (to appear). In this section, I will examine whether the presence of gender marking allows German speakers to make use of rare nouns more frequently. I present the results of three corpus studies showing that German gender marking coincides with an increase in the lexical richness of nouns cross-linguistically (compared to English) and intra-linguistically.

Since nouns are easier for users of a gendered language to produce, we expect to find a greater lexical richness of nouns in these languages. In a crude economic metaphor, suppose I have to pay a dollar for every bit of information in each word I utter, that nouns range in cost from $5 to $20, and that I am hesitant to pay more than $10 for any single word, though I'll do it if I have to. Grammatical gender systematically predicts every word, lowering its information; $3 is marked down from the cost of every word. Thus the former $13 words, which are rarer and more specific, are now $10, within my financial reach. Similarly, by facilitating noun recall, noun class can allow me to use a greater variety of nouns because it is easier for me to learn nouns. By bringing more words into the fold of easy produceability, grammatical gender should increase lexical richness of nouns.

### 3.1 Lexical richness of nouns

Say we compare English and German. Due to the systematic prediction of nouns by the grammatical gender markers, German speakers have easier access to a wider variety of nouns than English speakers. We thus expect greater lexical richness in German nouns than in English nouns. We can measure lexical richness using a simple type/token ratio, with sample size held constant. The higher the type/token ratio, the greater the lexical richness. Conveniently, type/token ratio is simply the inverse of average frequency, allowing for straightforward statistical testing: the lower the average frequency, the greater the lexical richness. Alongside lexical richness, I will examine the entropy of nouns in German and English, to see if a relationship exists between entropy reduction and increased lexical richness.

Word frequency distributions are difficult to handle using standard statistical methods because it is very difficult to find a measure of lexical richness that is not dependent on sample size (Baayen 2001: 24). To handle different sample sizes, word frequency distributions need to be characterized using LNRE (Large Number of Rare Events) models (Baayen 2008: 222). However, the samples available to me are too small for the subtleties of these models to be an issue, and the usual statistical comparisons are possible as long as sample size is strictly controlled.

**Method:** The NEGRA corpus was tagged for part of speech and compound noun status as in section 2. Compound nouns were removed. Counts of non-compound nouns immediately following definite articles were extracted for each article. Counts of non-compound nouns immediately following definite articles were then extracted from the New York Times Gigaword corpus of English, as parsed by the Stanford Parser (Klein &

Manning 2003). As a measure of the difficulty of nouns, I calculated the entropy of these nouns for English and German using the formula in section 2.3 above. I then calculated the conditional entropy given definite articles in German as in formula. As a measure of lexical richness, I computed a type/token ratio of lemmas in the German and English samples, with the English sample reduced to the same number of noun tokens as the German sample.

**Result:** The entropy of English nouns after definite articles is 10.17 bits. The entropy of German nouns after definite articles, with gender unknown, is 11.71 bits. The conditional entropy of German nouns after gendered definite articles is 10.55 bits.

The type/token ratio for German non-compound lemmas immediately following definite articles in NEGRA is 0.47, based on a sample of 16,640 nouns. The type/token ratio for English lemmas in the same context is 0.20. Type/token ratio is simply the inverse of average lemma frequency, so we can apply statistical tests measuring differences in means. The average frequency of the German lemmas is 2.12; the average frequency of a similar noun lemma in the NYT corpus is 4.93. A Welch two-sample t-test confirms the significance of the difference between these frequency distributions at $p < 0.001$.

**Discussion:** The entropy of English nouns that follow definite articles is about a bit and a half greater than the entropy of German nouns that follow definite articles, and close to the entropy of those German nouns given the gendered articles. The results of this experiment suggest that the Nominal Protection Device provided by noun class allows German speakers to use more informative nouns by using infrequent nouns more

frequently, that is, by selecting from a greater range of options for nouns. The increased entropy of nouns is offset by nominal protection.

The ever-present confound for this experiment in particular is genre. German newspaper writers may use more varied nouns than English newspaper writers, but this might not reflect facts about the spoken language but about cultural differences in newspaper style. In particular, German newspapers are notorious for a dense style (Clemens 1880), though the locus of complaint is usually complexity of syntactic structure rather than word choice. Unfortunately, no appropriate and large enough parsed corpus of spoken German exists to perform these calculations using spoken language data. Furthermore, in order to be sure that it is the NPD that affects word frequency distributions, we must examine data from a wider range of languages.

### 3.2 Language-internal evidence for lexical richness effects

We can also find language-internal evidence for the effects of a Nominal Protection Device on noun frequency distributions. Remember that no genders are distinguished in the plural in German. The NPD would be inoperative for plural nouns. Thus, we expect that the entropy of a plural noun out of context will be lower than that of a singular noun out of context.

**Method:** The entropy of plural and singular nouns in German was computed using the same method as in Section 3.1. Type/token ratios were collected. Lemmatized forms were used throughout.

**Results:** The entropy of singular nouns is 10.56 bits and the entropy of plural nouns is 8.92 bits. Singular nouns have a type/token ratio of 0.50 and an average

frequency of 2.01; plural nouns have a type/token ratio of 0.39 and an average frequency of 2.54, based on a sample of 2638 nouns. The difference in means is significant by a Welch Two Sample t-test at $p < 0.001$.

**Discussion:** The pattern is similar to the results of section 3.1. Singular nouns are more lexically rich than plural nouns, and on average would be harder to process out of context than plural nouns. The NPD allows singular nouns to have this distribution: the nouns are more readily available to speakers and more readily processed and remembered by hearers.

It is not yet clear, however, that this result is due to the effect of noun class. It might simply be that only a restricted subset of nouns can pluralize (for instance, abstract concepts very rarely pluralize). A cross-linguistic test is fitting here. In English, where plural nouns are no better or worse predicted than singular nouns, we should not find the same difference in lexical richness.

Unfortunately, we do find a similar difference: extracting nouns from the NYT corpus and lemmatizing them (as in section 3.1), we find a difference in average word frequency between singulars and plurals. Singulars have an average frequency of 2.47 and plurals have an average frequency of 2.84 with a sample of 2638 nouns; the difference is significant at $p=0.02$. However, note that the difference between English singulars and plurals is smaller in magnitude and less statistically distinct than the difference between German singulars and plurals.

### 3.3 Noun class and lexical poverty of adjectives

It is not immediately obvious that English is lexically impoverished in comparison to German, or that German communication is really more specific. It would not be a stretch to imagine that speakers maintain a similar level of specificity and information transfer cross-linguistically, especially within similar cultures. English is not without methods to accomplish specificity and the expression of rare concepts: the whole of context around the noun may be brought to bear in providing the additional information, or in providing a reduction in nominal entropy to facilitate the use of rare nouns. Such an approach may suggest, for instance, that English speakers make heavier use of greater variety of adjectives surrounding nouns in comparison to German speakers. Thus I predict that English is more lexically rich than German in its adjectives. German speakers have less use for adjectives because they can more easily produce rarer nouns.

There are two ways to look at this prediction. First, we can suppose that English speakers use more adjectives to compensate for the use of less varied nouns in communication; second, we expect that English speakers use more adjectives in order to better predict their nouns, so that adjectives facilitate the use of more difficult nouns. Essentially, the nominal prediction provided by noun class allows nouns to be more informative and less dependent on context; without the Nominal Protection Device, language users must make more use of extra-nominal context in facilitating noun use.

**Method**: All instances of Determiner-Adjective-Noun were extracted from the NEGRA corpus and the NYT corpus. All adjectives and nouns were lemmatized, removing the gender agreement markers in German. Controlling for token number, the lexical richness of adjectives and nouns in this context was compared.

**Result:** In a sample of 10,891 Determiner-Adjective-Noun triplets for German and English, the average frequency of a German adjective in this context is 6.66; the average frequency of an English adjective here is 4.08. The difference in means is significant at $p < 0.001$.

The average frequency of a German noun in this context is 2.26; the average frequency of an English noun is 3.36 ($p < 0.001$).

**Discussion**: As predicted, English adjectives are more lexically rich than German adjectives, though the reverse is true of nouns. To determine whether this is a result of the effects of grammatical gender on the informational structure of the noun phrase, or simply a difference in newspaper styles, the same calculation must be performed for other gendered and non-gendered languages. It may also be worthwhile to calculate the mutual information between adjectives and nouns in German, predicting less mutual information in German, because adjectives have less work to do in predicting nouns. Overall, gender allows German nouns to be more independent of their larger context, because German nouns are automatically better predicted in local context by the gender system.

### 3.4 Discussion

Noun class appears to facilitate the use of a broader range of nouns in a specific context in German as compared to English, but English speakers make up for the difference by using a richer set of adjectives. Germans can pack more information into their nouns without incurring processing difficulties, while English speakers spread that information out into the broader context, such as adjectives.

The results of this section are highly contingent on further cross-linguistic studies. Similar modern Germanic languages such as Dutch may provide one source, and historical Germanic languages with noun class, such as Old English, may provide a historical angle.

This is fine and good: but how is nominal protection achieved in the actual distribution of genders in the lexicon? In the next sections I will investigate the effects of semantics and discourse on gender assignment.

## 4. Semantic similarity and gender assignment

Another way in which noun class can help predict nouns is by predicting some aspect of their semantics. As Corbett (1991: 321) points out, semantically regular clusters are prevalent in noun class systems. Specifically in German, Zubin and Köpcke (1986) have investigated semantic regularities in German gender assignment. However, a systematic, corpus-based study of semantic regularities in gender systems has yet to appear.

Semantic regularity can be a double-edged sword for the purposes of the 'Nominal Protection Device.' On one hand, genders can provide useful semantic cues, but on the other hand, if gender assignment is too regular, then genders will be predicted by context and will provide no reduction in nominal entropy. But semantic regularity would only have that deleterious effect if gender assignment correlated with the semantic fields that frame large swaths of conversations.

In this section I present the results of a corpus study attempting to correlate semantic similarity with gender difference for pairs of German nouns, in an attempt to

provide the start of a systematic investigation of how clustering and diffusion coexist in German gender.

**Method:** In order to test whether a relationship exists between similarities in meaning and similarities in gender, a study was carried out using the Google Web 4-grams for German. Nouns from the CALLHOME German lexicon (Karins et al. 1997a) were each tested pairwise against each other noun for semantic similarity and for gender mismatch (coding the same gender as 0 and different, mismatching genders as 1). Semantic similarity was computed as the cosine of the angle between vectors of word counts in positions 3 words before and 3 words after each noun in the Google 4-grams. This method was adopted wholesale from Jurafsky & Martin (2009: 659 and 665). Preposition-article contractions were expanded, and all context words were stemmed to eliminate gender agreement morphology. A logistic regression model was fitted to predict gender mismatch on the basis of semantic similarity.

**Result:** A plot of the results is shown in Figure 3. As two nouns become more semantically similar, they are more likely to be in the same gender. Figure 3 presents the probability of gender mismatch for the noun pairs within each similarity group. The groups consist of noun pairs with similarities from 0 to 0.1, 0.1 to 0.2, etc. Similarity ratings range from 0 (no common context) to 1 (completely similar context). A least-squares regression line is added.

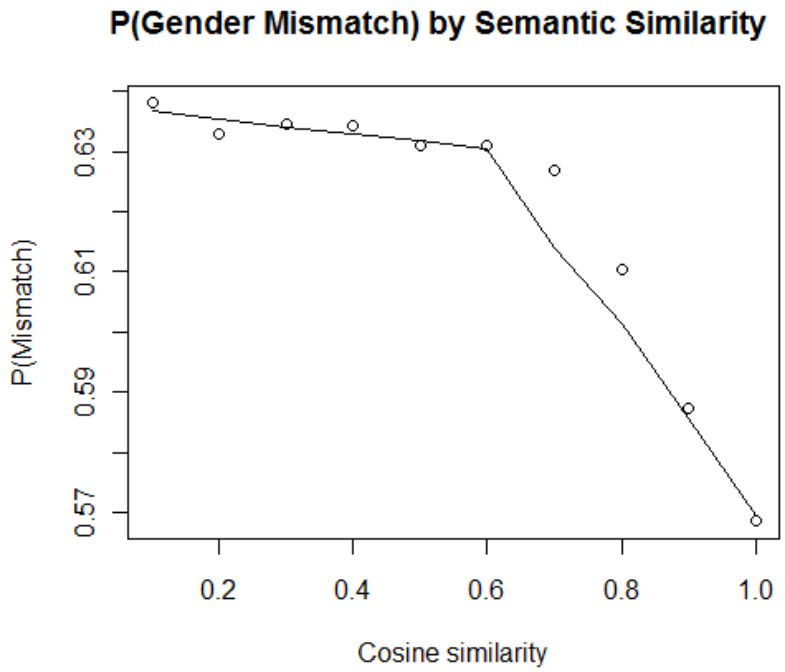


**Figure 3**. Probability of gender mismatch for pairs of nouns grouped by semantic similarity.

A logistic regression model confirms a negative correlation between cosine similarity and the gender mismatch for noun pairs at $p = 0.002$.

**Discussion:** That specific semantic fields may have one gender has already been demonstrated by Zubin and Köpcke (1986). The findings in this section show that those regularities, previously demonstrated only by example, have statistical validity in the whole lexicon. The close associations in the results come mostly from abstract nouns, which are nearly always feminine (although the negative correlation does persist with the feminine gender removed).

Given that some semantic fields will end up with similar genders due to the pressures of analogy, what we would want to find for the purposes of the NPD is that *frequent* semantic fields are the ones where gender assignment remains at the baseline frequencies, or is unusually even, since frequent semantic fields are the ones that would

cause problems most frequently if the 'Nominal Protection Device' failed in them. In small semantic fields, which are relatively unlikely to frame conversations, gender clustering can provide a useful semantic cue in predicting nouns.

## 5. Nominal Protection in discourse

The question remains as to whether German genders are distributed so that genders are not prematurely predicted in actual discourse. The question of gender distribution in the nouns used in actual conversations is particularly important for the use of nouns in predicting the referents of pronouns: if most nouns that would come up under a certain topic are of one gender, then the gender information in pronouns would provide no information. In this section I will present some work in progress about how the distribution of gender in the German lexicon promotes the use of the NPD in discourse.

If nouns are distributed to maintain heterogeneity in actual communicative situations, then nouns that co-occur should be in different genders as often as possible. The first study of this section tests that hypothesis directly.

### 5.1 Are nouns that co-occur are more likely to have different genders?

Each noun in the German CALLHOME corpus of telephone conversations was compared with every other noun pairwise on two points: (1) Are they of two different genders? (2) How many of the 100 CALLHOME conversations (Karins et al. 1997b) do they co-occur in? Logistic regression was performed in R (R Development Core Team 2005), predicting gender mismatch in a pair of nouns on the basis of how many conversations the nouns co-occur in.

**Result:** The results of the logistic regression are reproduced in Table 2, revealing a small but significant ($p = 0.04$) influence of co-ocurrence on gender mismatch in pairs of nouns. Co-occurance values ranged from 0 to 44, with values over 5 accounting for only 2% of co-occuring pairs. For this reason, co-occurence was simply coded as 0 (not co-occuring) or 1 (co-occuring at least once). For the dependent variable, gender mismatch is coded as 1 and gender match is coded as 0.

| Coefficient | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 0.59 | 0.002 | 372.8 | < 0.001 |
| Co-occurrence | 0.03 | 0.012 | 2.1 | 0.037 |

**Table 2**. Logistic regression model predicting gender mismatch from the number of co-occurrences in CALLHOME conversations.

If we simply code whether or not two nouns co-occur at all, ignoring how many conversations they co-occur in, the effect remains in similar strength. Figure 4 demonstrates how, for each pair of nouns, the probability of gender mismatch is larger when those two nouns co-occur in one or more conversations in the CALLHOME corpus.

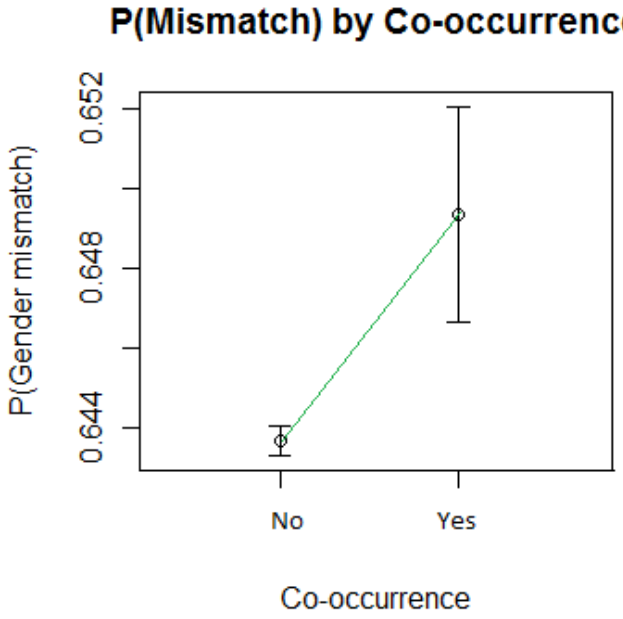


**Figure 4.** Probability of gender mismatch noun pairs depending on whether they co-occur in a conversation. 95% confidence intervals are displayed.

**Discussion:** As the number of conversations with a co-occurrence of two nouns increases, the chances of those two nouns having different genders increases. This configuration is a step in the right direction: nouns that co-occur are more likely to be confused in anaphoric pronominal reference, and co-occurrence is a diagnostic for nouns being common in the discussion of a certain topic. The results of this section seem to indicate that, while genders cluster in certain semantic subfields, they remain heterogeneous in the larger fields that frame conversations.

### 5.2 Frequency and gender distribution

One way to achieve a distribution of genders that optimizes the variety of nouns in actual discourse is to distribute genders most evenly in frequent nouns, since the more frequent a noun is, obviously the more likely it is to co-occur with any other noun. These

frequent nouns are more likely to occur in any conversation and contribute to variety of noun classes. To study this possibility, I examined to what extent frequency could account for the results of section 5.1.

**Method:** Frequency information was gathered from the CALLHOME corpus and the sum of the frequencies of each noun pair was incorporated into the logistic regression model reported above.

**Result:** The main effect on gender mismatch for German noun pairs is eliminated when we control for frequency. Co-occurrence now has an insignificant negative effect, while frequency has a main effect of similar size and direction to the effect of co-occurrence in section 5.1 The summed frequency was logged in order to account for the logarithmic distribution of word frequencies described by Zipf's Law (Zipf 1935). The results of the logistic regression are displayed in Table 3.

| Coefficient | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 0.552 | 0.003 | 173.6 | < 0.001 |
| Co-occurrence | -0.021 | 0.012 | 0.4 | 0.12 |
| Log sum frequency | 0.027 | 0.002 | 17.3 | < 0.001 |

**Table 3**. Logistic regression model predicting gender mismatch from the number of co-occurrences in CALLHOME conversations and frequency.

**Discussion:** Frequency accounts for the gender distribution indicated in section 5.1. The negative effect of co-occurrence when controlling for frequency, if shown to be significant with a larger dataset, would perhaps reveal an effect of semantic similarity in causing similarity of gender assignment.

### 5.3 Discussion

Together with the results of section 4, there is some evidence here that genders are distributed so that they are not predicted in discourse context, thus mitigating their usefulness in nominal prediction. While semantic regularities exist, the distribution of genders according to frequency is such that frequent nouns inject a variety of gender into discourse. Thus noun class remains useful in protecting nouns by predicting them.

## 6. Summary and Conclusion

### 6.1 Summary and interpretation of results

I presented the results of three studies focusing on three aspects of the Nominal Protection Device. First, noun classes reduce nominal entropy, which has effects in German versus English speakers' use of nouns. From a semantic angle, I found that noun classes contain semantic clusters, which can be useful in predicting the semantics of nouns in specific semantic subfields. Finally, noun classes are distributed so that nouns that co-occur are more likely than chance to be in different genders: noun class is thus useful in predicting nouns in discourse.

### 6.2 Suggestions for further research

A number of proposals concrete and general emerge from the present work. The study of the effect of noun class on other, probabilistic elements of language structure, such as noun frequency and pronoun use, must be extended to a broad range of languages to determine if an effect does exist. If historical corpora of sufficient size exist, one could trace the evolution of noun class systems as part of the larger communicative system of language: can changes in how information is expressed bolster or weaken a noun class

system? Similar approaches can be explored language-internally: do the specific different genders of German, with their differing degrees of information about the following noun, cause speakers to use nouns differently, or are speakers affected only by the reduction in uncertainty provided by the system as a whole?

The study of the relation of noun classes to semantic classes must be studied in more depth. For noun class to work as the 'Nominal Protection Device', clustering and diffusion must coexist in the lexicon. I suggested in section 4 that more frequent semantic fields may especially avoid subcategorizing for certain genders, but other hypotheses are possible. The relationship between noun class and extra-nominal context remains a fundamental issue: how are noun classes distributed to prevent their prediction in context while still providing semantic cues to a noun?

### 6.3 Conclusion: Learning, Form, and Function

I would not consider the hypothesis of noun class as a Nominal Protection Device proved by the studies in this paper. But I hope that the results of the specific studies and the general information-theoretic approach to noun class prove helpful or interesting to other researchers.

Nevertheless, I believe the results of the co-occurrence study and the results cited in sections 1.2.1-1.2.4 (on the functions of noun class) give reason to continue to advance the NPD hypothesis. It is difficult to maintain that noun class really is the product of the "nonsensibleness of the human race." One may claim that the reason generation after generation of children learn noun class systems is nothing other than that their learning

mechanisms pick up on everything, even redundant agreement markers. But the aspects that make noun class helpful are the aspects that make noun class learnable.

As Inbal & Ramscar (to appear) show, noun class marking helps language users learn nouns, and the way in which it helps users learn nouns makes noun class learning essential. Noun classes, in the form of prefixed articles or referring pronouns, provide a set of linguistic cues to all nouns. That fact is captured in the description of noun classes as reducing nominal entropy: we see that noun classes bear some of the informational load of nouns. If noun class markers no longer brought about a reduction in uncertainty about nouns—if the class markers were fully predicted by extra-nominal context, or if noun class distribution became unequal, the scenario in section 2—then they would no longer be learned due to blocking.

The reason for this is apparent in the theory of language learning presented in Ramscar et al. (2010), in which children learn labels as cues for features, or features as cues for labels. The crucial detail here is that learning is directed: children learn salient features only as cues for labels, and similarly would not learn labels unless they were cues to features. In the case of noun class, children are learning noun class as a cue for nouns, and nouns as cues to noun class markers. If the noun class no longer told the child anything about the noun, it would not be a cue to the noun, and would not be learned. The properties that make noun class markers useful in reducing nominal uncertainty and in providing reference tracking are the same properties that ensure that noun class is learned. Thus, by finding information-theoretically motivated functions for noun class, we can provide a tentative explanation for the existence and persistence of noun class systems.

More generally, the work I present here can be considered part of the line of research exemplified by Piantadosi et al. (2009), who advance the "Communicative Lexicon Hypothesis," the idea that the lexicon of language is structured in a way such as to facilitate the effective use of language by reducing redundancy and enhancing informative elements. And the more functional a linguistic system is in an information-theoretic sense, the more strongly that system will be learned. In this sense work on noun class and on the Communicative Lexicon Hypothesis in general can bridge the gap between function and form in linguistic explanation.

And yet on a less theoretical scale, I hope to give some hope to second language learners, teachers, and linguists who encounter these byzantine gender systems. Gender is not arbitrary and it is not useless: it is in fact an ingenious system that provides language users with a set of keys to the hardest-to-access parts of language.

Acknowledgments

This work would have been impossible without huge help and patience from my advisor Dan Jurafsky, conceptual help from Michael Ramscar, and statistical help from Victor Kuperman. Thanks are also due to Rachel Antonsen for editing and María Helga Guðmundsdóttir for assistance with MATLAB. However all faults are due only to me.

Works Cited


Arnon, Inbal & Michael Ramscar (under review) Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned.

Baayen, R. Harald (2008). *Analyzing linguistic data*. New York: Cambridge University Press.

Baayen, R. Harald (2001) *Word frequency distributions*. Boston: Kluwer Academic Publishers.

Barlow, G. M. (1992). *A Situated Theory of Agreement*. New York, Garland Publishing, 328. 1992

Clark, H. H.**,** & T. Wasow (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37, 201-242.

Claudi, Ulrike (1985). *Zur Entstehung von Genussystemen: Überlegungen zu einigen theoretischen Aspekten, verbunden mit einer Fallstudie des Zande: Mit einer Bibliographie und einer Karte.* Hamburg: Buske.

Clemens, Samuel (1880). The awful German language. Available online at http://www.crossmyt.com/hc/linghebr/awfgrmlg.html.

Cohen Priva, Uriel (2008). Using Information Content to Predict Phone Deletion, *Proceedings of the 27th West Coast Conference on Formal Linguistics*, ed. Natasha Abner and Jason Bishop, pp 90-98. Cascadilla Proceedings Project: Somerville, MA.

Coombs, C. H., R. M. Dawes, & A. Tversky (1970). *Mathematical psychology: An elementary introduction*. Englewood Cliffs, NJ: Prentice-Hall.

Corbett, Greville G. (2008). Systems of Gender Assignment. In: Haspelmath, Martin & Dryer, Matthew S. & Gil, David & Comrie, Bernard (eds.) *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library, chapter 32. Available online at http://wals.info/feature/32 Accessed on 2009-12-07.

Corbett, Greville G. (1991). *Gender*. Melbourne: Cambridge University Press, 1991.

Corbett, Greville G. (1979). The Agreement Hierarchy. *Journal of Linguistics* 15, 203 -224.

Dixon, R. M. W. (1968). Noun classes. *Lingua* 21: 104-25.

Eisenberg, Peter (2006). *Grundriss der deutschen Grammatik*. Metzler Verlag.

Genzel, D. and E. Charniak (2002). Entropy rate constancy in text. In *Proceedings of ACL-02*.

Gibson, Edward (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1-76.

Greenberg, Joseph H. (1963). Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements." In: Joseph H. Greenberg (ed.) *Universals of Language*. London: MIT Press: 73-113.

Heath, Jeffrey (1975). Some functional relationships in grammar. *Language* 51: 89-104.

Itti, L. & P. Baldi (2005). Bayesian surprise attracts human attention. In *Advances in neural information processing systems*.

Jaeger, T. Florian (2009 ms). Redundancy and Reduction: Speakers Manage Syntactic Information Density.

Klein, Dan & Christopher Manning (2003). Fast Exact Inference with a Factored Model for Natural Language Parsing. In *Advances in Neural Information Processing Systems* 15 (NIPS 2002), Cambridge, MA: MIT Press, pp. 3-10.

Koval', A.I. (1979). O značenii morfologičeskogo pokazatelja klassa v fula. In N. V. Oxotina (ed.) *Morfonologija i morfologija klassov slov v jazykax Afriki*, 5-100. Moscow: Nauka.

Köpcke, Klaus-Michael (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache.* Tübingen: Niemeyer (Linguistische Arbeiten 122).

Karins, Krisjanis, Robert MacIntyre, Monika Brandmair, Susanne Lauscher & Cynthia McLemore (1997a). *CALLHOME German Lexicon*. Linguistic Data Consortium, Philadelphia.

Karins, Krisjanis, Robert MacIntyre, Monika Brandmair, Susanne Lauscher & Cynthia McLemore (1997b). *CALLHOME German Transcripts*. Linguistic Data Consortium, Philadelphia.

Lakoff, George (1986). Classifiers as a reflection of mind. In *Typological Studies in Language 7: Noun Classes and Categorization* (ed. Colette Craig), pp. 13-51.

Lezius, Wolfgang, Reinhard Rapp, Manfred Wettler (1998). A Freely Available Morphological Analyzer, Disambiguator, and Context Sensitive Lemmatizer for

German. In *Proceedings of the COLING-ACL 1998*: 743-747.

Maratsos, M. P. (1979). How to get from words to sentences. In D. Aaronson & R. Rieber (eds.), *Perspectives in psycholinguistics*. Hillsdale, N.J.: Erlbaum, 1979.

McDonald, S. A. & R. C. Shillcock (2003). Eye movements reveal the on-line computation of lexical probabilities. *Psychological Science, 14*, 648-652.

Piantadosi, Steven, Hal Tily, & Edward Gibson (2009). The communicative lexicon hypothesis. In N. Taatgen & H. van Rijn (eds.), *Proceedings of CogSci 2009*.

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramscar, Michael, Dan Yarlett, M. Dye, K. Denny, & K. Thorpe. (2010) Feature-Label-Order effects and their implications for symbolic learning. Cognitive Science.

Robinson, Orrin. (to appear). Does Sex breed Gender?: Pronominal Reference in the Grimms' Fairy Tales," to appear in *Marvels and Tales: Journal of Fairy-Tale Studies (special issue on New Approaches to Fairy Tale Scholarship)*

Shannon, Claude E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, July and October.

Skut, W., B. Krenn, T. Brants., & H. Uszkoreit. (1997). An annotation scheme for free word order languages. In *Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP)*, Morgan Kaufmann, San Francisco.

Tucker, G. R., W. Lambert, A. Rigualt, & N. Segalowitz. 1968. A psychological investigation of French speakers' skills with grammatical gender. JVL & VB, 7/2, 312-316.

Vigliocco, G., T. Antonini and M. F. Garrett (1997). Grammatical Gender Is on the Tip of Italian Tongues. *Psychological Science*, 8/4, 314-317.

Zipf, George K. (1935). The psycho-biology of language: An introduction to dynamic philology. Houghton Mifflin.

Zubin, D. A. & Klaus-Michael Köpcke (1986). Gender and folk taxonomy: the indexical relationship between grammatical and lexical categorization. In C. G. Craig (ed.)

*Noun Classes and Categorization: Proceedings of a Symposium on Categorization and Noun Classification, Eugene, Oregon, October 1983*. (Typological Studies in language, 7), 139-80. Amsterdam: Benjamins.