Paradigmatic structure in speech production

R.H. Baayen[1], W.M.J. Levelt[2], R. Schreuder[3], and M. Ernestus[3]
[1]University of Alberta, [2]Max Planck Institute for Psycholinguistics, and [3]Radboud University

1 Introduction

In many of the world's languages, nouns are inflected for number. In general, the singular is simpler than the plural, both with respect to form, and with respect to meaning. For instance, the English singular *nose* consists of just the bare stem *nose*, while the plural is created from the singular by adding the suffix *-s*. This difference in formal complexity is mirrored in the complexity of the corresponding semantics, with the singular typically refering to one and the plural to two or more instances of the noun's referent. Using the terminology of structuralist linguistics, the singular is the unmarked, and the plural the marked form.

While most nouns denote objects that typically occur singly, there are also nouns with referents that typically occur in pairs or groups. In English, nouns such as *eye* and *sheep* come to mind. Some languages have a special inflectional form, the dual, for referring to paired objects. In yet other languages, nouns referring to objects that typically occur in pairs or groups are referred to by a simple form that carries a plural meaning. To express the singular, a singulative suffix has to be added (Dimmendaal 1987). In languages such as English, plurals for nouns with referents for which dual or plural occurrence is more natural than singular occurrence were described by Tiersma (1982) as locally unmarked. Thus, *nose* is an unmarked singular, and *eyes* is a locally unmarked plural.

Local markedness is reflected in the relative frequencies of the singular and plural form in the number paradigm. Given the summed frequency of the singular and plural form, henceforth Lexeme Frequency (using 'lexeme' in the sense of Aronoff (1994)), locally unmarked plurals tend to be more probable than their corresponding singulars. Conversely, globally marked plurals tend to be less probable than their singulars. In what follows, we will refer to nouns for which the singular is more frequent than the plural as singular-dominant, and to nouns for which the plural is more frequent than the singular as plural-dominant.

A first study on the comprehension of singular-dominant and plural-dominant nouns in Dutch (Baayen *et al.* 1997b) made use of visual lexical decision to probe written comprehension. A factorial design contrasted Dominance (singular-dominant versus plural-dominant), Number (singular versus plural), and Lexeme Frequency (high versus low). Figure 1 illustrates the pattern of results obtained. Low-frequency lexemes (in grey) elicited longer latencies than high-frequency lexemes (in black), plural-dominant plurals elicited shorter latencies than singular-dominant plurals, and singulars elicited similar

latencies irrespective of Dominance. The authors took this to indicate that the unmarked singular inherited the frequency of the plural form. The crucial predictor for the marked plural, by contrast, was taken to be its own frequency.
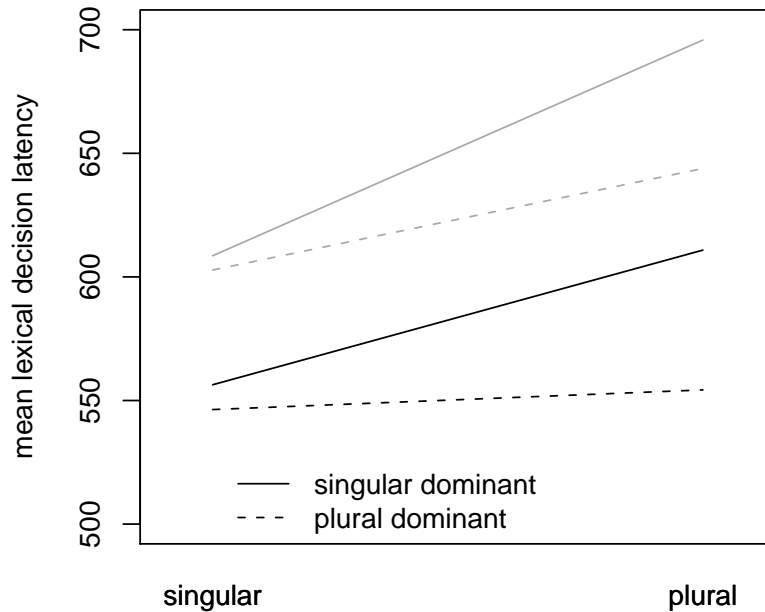
Figure 1: Interaction plot for visual lexical decision latencies for singular and plural nouns in Dutch, cross-classified by Dominance and Lexeme Frequency. Grey lines represent nouns with low lexeme frequency, black lines represent nouns with high lexeme frequency.

A very similar pattern of results was subsequently observed for auditory comprehension (Baayen *et al.* 2003). Effects of Lexeme Frequency and Dominance were also observed for Italian (Baayen *et al.* 1997a) and French (New *et al.* 2004). For English, however, New and colleagues observed an effect of Dominance, but response latencies for English singulars were not straightforwardly predictable from their lexeme frequency: singular-dominant singulars elicited shorter visual lexical decision latencies than plural-dominant singulars (see also Sereno & Jongman 1997).

The main goal of the present study is to trace the consequences of local and global markedness for the production of singular and plural nouns, again using Dutch as language of investigation, thereby complementing the body of experiments on the comprehension of number inflection. Given the dual mechanism model (Pinker 1997; Pinker 1999) and the WEAVER model (Levelt

et al. 1999), the prediction is that no independent frequency effects should be observed for plurals in speech production. Dutch plural formation is regular, so all that is needed to produce a plural form is access to its constituent morphemes, the stem and the plural suffix (Baayen *et al.* 2002; Keuleers *et al.* 2007). Accessing two morphemes instead of one may lead to an effect of Number (with access to two morphemes requiring more processing resources and hence leading to longer latencies), accessing the stem may lead to an effect of lexeme frequency (see, e.g., Jescheniak & Levelt 1994). However, Stemberger & MacWhinney (1986); Stemberger & MacWhinney (1988) reported that high-frequency regularly inflected forms are less prone to speech errors. Their results suggest that intensively used regular inflections might develop their own form representations, which would then protect against speech errors.

We studied the processing of inflection in speech production with a series of (thus far unpublished) experiments, the first three of which were run in the nineties[1], and the last of which was run two years ago. Experiment 1, using the picture naming paradigm, was designed to address the processing of plural inflection in production by means of the same factorial contrasts used by Baayen *et al.* (1997b).

The effect of Dominance that emerged from Experiment 1 indicated, surprisingly, that plural-dominance apparently leads to slowed picture naming latencies for both the singular and plural forms. Experiments 2 and 3 were designed to rule out that this reversed effect of dominance arises at the level of articulation or at the level of picture interpretation and conceptualization. These two experiments allow us to establish that the effect of dominance takes place during lexical access. Experiment 4 is a replication study of the first experiment using new materials and a new design that in addition addresses the interpretation of pictures with exactly two versus pictures with more than two objects.

The present paper also has two subsidiary goals. The first of these is to illustrate how central concepts from information theory can help us understand aspects of lexical processing (see also Milin *et al.* 2008b; Kuperman *et al.* 2008). A second, methodological, goal is to illustrate that dichotomization of continuous predictors can be harmful and stand in the way of a proper understanding of experimental data. Various statistical studies (Cohen 1983; Maxwell & Delaney 1993; MacCallum *et al.* 2002) have warned against dichotomization of quantitative variables. We were not aware of these studies when more than 10 years ago we designed the experiments reported in Baayen *et al.* (1997b) and Experiments 1–3 of the present study. In what follows, we will show that factors that dichotomize an underlying numeric predictor should be replaced by that numerical predictor itself in the statistical analysis. It turns out that this is crucial for understanding Experiments 1 and 4, and also leads to a simpler and superior model for understanding the visual lexical decision data discussed in Baayen *et al.* (1997b).

---

[1]We are indebted to Allette Haveman and Ger Desserjé for their assistence with the preparation and running these experiments.

2 Experiment 1: Picture Naming
2.1 Method

Materials Sixty-four picturable nouns were selected for presentation in two conditions. In the singular condition, a simple line drawing of a single typical instance of the noun's referent was shown. In the plural condition, two slightly smaller versions of the same picture (reduced to 70% of the original size) were shown side by side (see
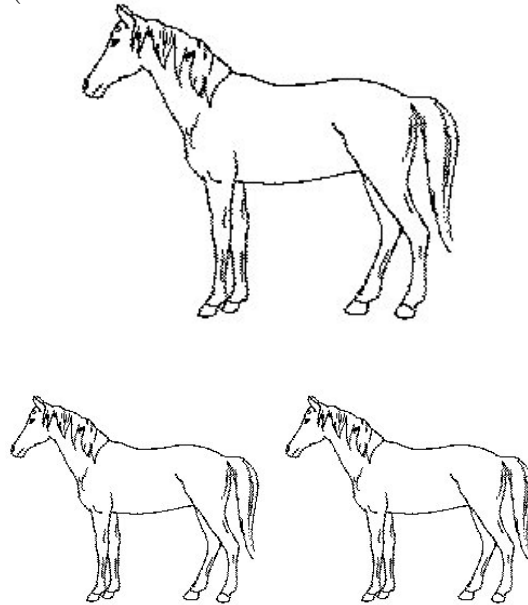
Figure 2: An example of the kind of line drawings used in Experiment 1: singulars (above) and plurals (below).

Thirty-two nouns had a high lexeme frequency (mean 2326 per 42 million), and thirty-two nouns had a low lexeme frequency (393 per 42 million). Henceforth, we will refer to this factorial contrast as Lexeme Frequency. The lexeme frequency of a noun was estimated using the 'lemma' frequency information in the CELEX lexical database (Baayen *et al.* 1995), and is equal to the summed frequency of the lexeme's inflectional variants. For each of the levels of Lexeme Frequency, 16 nouns were singular dominant, and 16 were plural dominant. We refer to this factorial contrast as Dominance. For singular-dominant nouns, the singular was on average 5.3 times as frequent as the plural. For plural-dominant nouns, the plural was on average 1.8 times as frequent as the singular. The words in the two Lexeme Frequency conditions were matched for Dominance.

All noun singulars were monosyllabic words, three to five phonemes in length with a mean length of four phonemes in each group. All require resyllabification after suffixation with the plural suffix (e.g., *worm*, 'worm', *wor$men*, 'worms'). Five plural forms (two plural dominant, three singular dominant) further differed with respect to the voicing of the word-final obstruent of the

singular, as in *laars*, 'boot', *laarzen*, 'boots'. The probabilistic grammar governing these changes is described in Ernestus & Baayen (2003). For a list of the materials, the reader is referred to Appendix A.

We divided the sixty-four nouns into two lists of 32 lexemes, such that the lexeme frequencies as well as the ratios of singular to plural frequencies were approximately matched across the two groups. From each of these two lists, we created two versions (A and B) that differed with respect to Number. List version A contained 16 plural-dominant singulars and 16 singular-dominant plurals. List version B contained the remaining 16 plural-dominant plurals and the remaining 16 singular-dominant singulars. (Both A and B list versions were balanced with respect to Lexeme Frequency and the Dominance of the lexemes.) A given subject was therefore exposed to a total of 32 lexemes, and was asked to name only (locally) marked forms (list A) or (locally) unmarked forms (list B). This between subjects design was chosen in order to rule out that effects might be artificially induced by having both marked and unmarked forms together in a block. We also ran the same experiment without blocking by dominance. Results were indistinguishable from the results reported here. This allows us to conclude that our results do not depend on list composition.

The order of the items in the four lists was pseudo-randomized such that not more than five singulars and not more than five plurals occurred in sequence. We presented a given list three times to a given subject. We will refer to this variable as Exposure. The experiment was preceeded by a practice session with line drawings of 25 singular and 25 plural objects, none of which appeared later in the experiment.

Subjects Thirty-two subjects, students at the university of Nijmegen, were paid to participate in the experiment. All had normal or corrected-to-normal vision and no known speech impairment. Half of the subjects were presented with list version A, the other half with list version B.

Procedure Subjects were tested in groups of two, each in noise-attenuated experimentation booths. Before the experiment, we took our subjects through a picture book in which the line drawings of singular and plural pairs were presented together on a page. The singular was shown on the upper half, and the plural on the lower half of the page, with the singular or plural word printed underneath the corresponding drawing. In this way, we ensured that our subjects would understand the pictures and would name them appropriately.

Each trial consisted of a fixation mark (asterisk) in the middle of the screen during 200 ms, followed after 600 ms by the picture centered at the same position. The pictures were presented on Nec Multisync color monitors in white on a dark background and remained on the screen until a subject responded by naming the picture, or disappeared after the time-out of 2000 ms if no response was given. A new trial was initiated 1500 ms after response or time-out. Naming latencies were measured from picture onset. Four pauses were included in the experiment, one following the practice session, and three pauses of 30 seconds between four blocks of 48 trails. The total duration of the experiment was approximately 30 minutes.

2.2 Results and discussion

Subjects performed this experiment with a high degree of accuracy, with an error rate of 5.8%. We classified a trial as an error when the response exceeded the timeout of 2000 ms, or when there was a voicekey or naming error. Before analysing the naming latencies, we first removed data points for which an incorrect response had been recorded from the data set. Inspection of the order statistics of the remaining naming latencies revealed marked non-normality, most of which was eliminated by removing extreme outlier data points (RT ≤ 400 or RT ≥ 1200) and by logarithmically transforming the latencies. These cutoff points were identified by visual inspection, and corresponded to a lower bound of 2.2 standard deviations below the mean, and 2.7 standard deviations above the mean. (Symmetrical cut-off points around the mean would either have left clear outliers in the data set at the lower end of the distribution, or would have removed too many non-outlier data points at the higher end of the distribution.) Table 1 summarizes the mean response latencies, and Figure 3 presents the corresponding interaction plot.

	low lexeme frequency	high lexeme frequency
singular dominant singulars	636	608
singular dominant plurals	678	635
plural dominant singulars	678	634
plural dominant plurals	692	653

Table 1: Mean naming latencies for Experiment 1.

The interaction plot shows that, unsurprisingly, the stimuli from the low lexeme frequency condition (dashed lines) elicited longer naming latencies than the stimuli from the high lexeme frequency condition (solid lines). The fact that plurals (right) elicited longer latencies than the singulars (left) is also as expected: the plurals have a more complex phonological structure with two syllables instead of one. What is surprising is that the naming latencies for plural-dominant nouns were longer than those of singular-dominant nouns, irrespective of whether the noun to be named was a singular or a plural. In comprehension, high-frequency plurals have a processing *advantage* compared to low-frequency plurals. In this production experiment, by contrast, plural dominance gives rise to a processing *disadvantage* that extends to both the singular and the plural form.

Table 2 presents the contrast coefficients for Number, Dominance and Lexeme Frequency as estimated by a linear mixed-effects model with Subject and Lexeme as crossed random effects (Bates 2005; Baayen *et al.* 2008; Baayen 2008). Interactions did not reach significance. List was initially included as random effect, but its variance estimate turned out to be effectively zero, and was therefore removed from the model. The main effects of Number and Lexeme Frequency received unequivocal support: both the *p*-value based on the *t*-distribution (with the upper bound for the degrees of freedom, see Baayen *et al.* 2008) and the more conservative *p*-value based on the posterior dis-
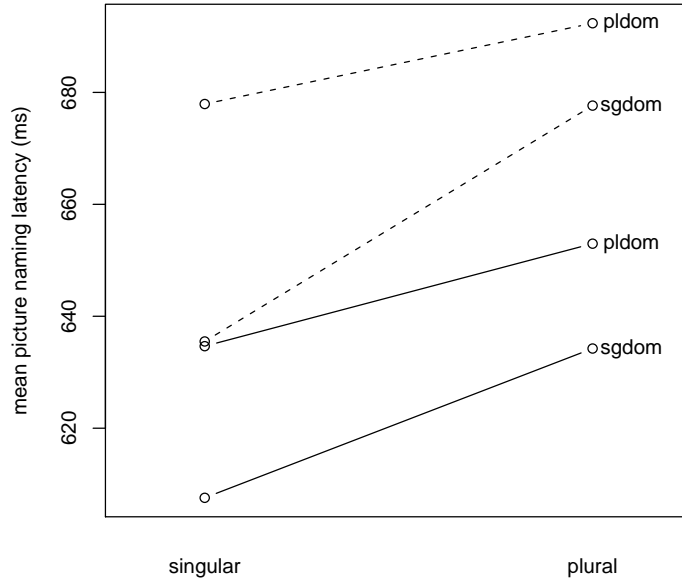
Figure 3: Interaction plot for Experiment 1. Solid lines connect the high lexeme frequency conditions, dashed lines represent the low lexeme frequency condition. In the plot, plural-dominance is denoted by *pldom*, and singular-dominance by *sgdom*.

	Estimate	lower HPD	upper HPD	p (MCMC)	p (t)
Intercept	6.5916	6.5461	6.6399	0.0001	0.0000
Number sg	-0.0363	-0.0446	-0.0274	0.0001	0.0000
Dominance sgdom	-0.0404	-0.0830	0.0024	0.0706	0.0298
Lexeme Frequency low	0.0642	0.0239	0.1098	0.0052	0.0006
Exposure	-0.0566	-0.0618	-0.0512	0.0001	0.0000

Table 2: Analysis of variance of Experiment 1 (Picture Naming). Estimate: estimated beta weights (using contrast coding for factors). Lower, upper HPD: lower and upper bounds of the 95% Highest Posterior Density intervals for the coefficients; p (MCMC) denotes the corresponding Markov chain Monte Carlo *p*-value; p (t) denotes the *p*-value based on the *t*-distribution; sg: singular; sgdom: singular dominant.

tribution of the parameters according to 10000 Markov chain Monte Carlo (MCMC) samples were well below 0.05. Although the effect of Dominance was supported by the p-value based on the t-distribution, it did not reach full significance when evaluated on the basis of the MCMC samples.

In this factorial analysis, Lexeme Frequency and Dominance are factors that dichotomize underlying gradient predictors. Various statistical studies (Cohen 1983; Maxwell & Delaney 1993; MacCallum *et al.* 2002) have warned against dichotomization of quantitative variables. We therefore examined the data of Experiment 1 in further detail, replacing the factor Lexeme Frequency by the actual frequency of the lexeme. Furthermore, we replaced Dominance by an information theoretic measure, Entropy, that emerged as a significant predictor in comprehension studies (Moscoso del Prado Martín *et al.* 2004; Baayen *et al.* 2006).

Let i range over the singular and plural forms of a lexeme, and let p_i denote the probability of a singular (or plural) form given the lexeme. We estimate p_i by the relative frequency of the form given its number paradigm. For a singular with frequency 75 and a plural with frequency 25, the respective probabilities are 0.75 and 0.25 respectively. The entropy (Shannon & Weaver (1949)) of the number paradigm can now be introduced as

$$(1) \quad H = -\sum_i p_i \log_2(p_i).$$

The Inflectional Entropy H quantifies the average amount of information in the number paradigm expressed in bits. The entropy is greatest when the probabilities of the inflectional variants are uniformly distributed. Since the probabilities of the singular and plural form, $P(sg)$ and $P(pl)$, sum to unity, the graph of H is symmetrical around $P(sg) = P(pl) = 0.5$, the value for which it reaches its maximum, 1, as illustrated in Figure 4. Note that the entropy of, for instance, a noun with $P(sg) = 0.7$ and $P(pl) = 0.3$ is identical to the entropy of a noun with $P(sg) = 0.3$ and $P(pl) = 0.7$.

Using (1), we calculated the inflectional entropy for each noun, where it is crucial to keep in mind that the inflectional entropy is identical for the singular and the plural form of a given noun. The mean inflectional entropy was 0.498 for the singular dominant nouns, and 0.606 for the plural dominant nouns ($t(56.098) = -3.8092, p = 0.0003$).

The reason that these values differ for the two sets of nouns turns out to be an imbalance in the extent to which the frequency of one form exceeds the frequency of the other form. For globally unmarked nouns, the singular tends to be substantially more frequent than the plural. For locally unmarked nouns, although the plural is more frequent than the singular, the frequency imbalance is muted compared to the globally unmarked nouns. The extent to which the frequency of a locally unmarked plural exceeds the frequency of its corresponding singular tends to be substantially reduced compared to the extent to which the frequency of an unmarked singular tends to exceed the frequency of its plural. As a consequence, the frequencies of the singular and plural forms tend to be much more similar for plural-dominant nouns, and this
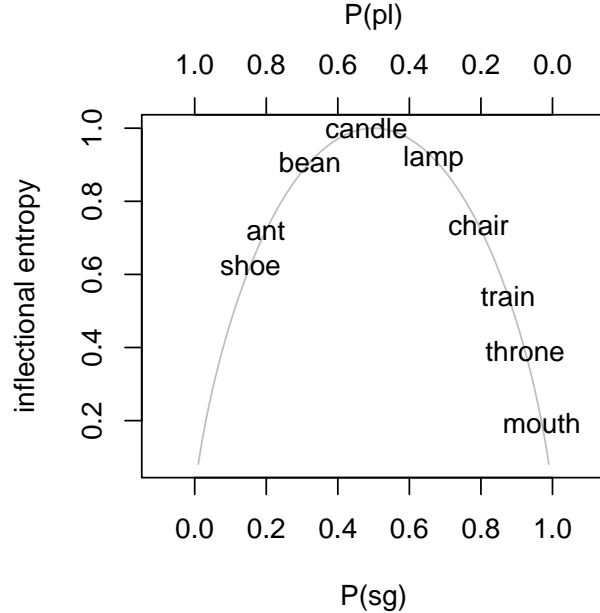
Figure 4: Inflectional Entropy as a function of the probability of the singular (or the plural), with selected examples of nouns in Experiment 1.

is why plural-dominant nouns tend to have higher inflectional entropies. This is not an artifact of our sample, but reflects the dominance relations in the lexicon. Although plural-dominant plurals are locally unmarked in the sense of Tiersma (1982), the markedness reversal for plurals never reaches the extremes characteristic of singular dominance. This is illustrated graphically in Figure 4: plural-dominant lexemes are found to the left with less extreme values on the horizontal axes than the singular-dominant lexemes in the right-hand side of the graph.

When Entropy is added as a predictor to the factorial model, it emerges as significant to the exclusion of Dominance. Note that because the entropy is the same for the singular and plural form of a lexeme, it will effect both inflectional variants in the same way, which was exactly what the main effect of the factor Dominance achieved. However, with entropy we have a more powerful predictor than the original factor.

In what follows, we introduce a more complete regression model fitted to the data of Experiment 1. In addition to a noun's Lexeme Frequency, Number, and Inflectional Entropy, we included Trial (how far a subject had progressed in the experiment, an experimental control variable), Exposure, Length (in phonemes), and the noun's number of meanings (Jastrzembski (1981)), gauged by means of a count of the synonym sets (synsets) in the WordNet database (Miller 1990; Fellbaum 1998; Baayen *et al.* 2006). We included number of meanings as a covariate because we had previously observed it to be predictive both in visual lexical decision and word naming (Baayen *et al.* (2006)), and wanted to ascertain whether it would emerge as well in a naming task that is

conceptually driven (instead of driven by the visual input as in word naming). Furthermore, by including the synset count, we obtain some control over the words' meanings. If the effect of dominance is not confounded with semantic ambiguity (which, as a lexemic property, could affect the singular and plural form in the same way), then it should remain significant in a model that also includes the synset count as a predictor.

|	Estimate	lower HPD	upper HPD	p (MCMC)	p (t)
Intercept	6.5055	6.3800	6.6654	0.0001	0.0000
Frequency	-0.0219	-0.0339	-0.0095	0.0004	0.0004
Trial (linear)	0.0003	-0.0255	0.0250	0.9918	0.0235
Trial (quadratic)	0.0000	0.0000	0.0000	0.0008	0.0004
Synsets (linear)	-0.0202	-0.0456	0.0038	0.1274	0.1107
Synsets (quadratic)	0.0032	0.0000	0.0060	0.0456	0.0355
Length	0.0277	0.0069	0.0432	0.0076	0.0020
H	0.1685	0.0292	0.2706	0.0144	0.0053
Exposure	-0.0519	-0.0588	-0.0442	0.0001	0.0000
RE	0.0961	0.0370	0.1434	0.0002	0.0003
Exposure:RE	-0.0308	-0.0444	-0.0151	0.0001	0.0000

Table 3: Regression analysis of Experiment 1 (Picture naming). Frequency: lexeme frequency, Trial: position in the experimental list, Synsets: number of synsets in WordNet, Length: length in phonemes, H: inflectional entropy, Exposure: count of repeated presentation, RE: relative entropy.

A final measure that we included is Relative Entropy (RE in Table 3). Relative entropy quantifies the extent to which the probability distribution of a particular noun diverges from the corresponding probability distribution of the class of nouns. More specifically, let P denote the probability distribution of a given noun's number paradigm. For the plural-dominant noun *laars* ('boot'), $P = \{183/(183 + 861), 861/(183 + 861)\} = \{0.175, 0.825\}$. Let Q denote the probability distribution of the inflectional 'class', i.e., the set of all singular and all plural nouns. For Dutch, this reference distribution $Q = \{0.748, 0.252\}$. With i ranging over the two probabilities associated with singulars and plurals in the P and Q distributions, the relative entropy is defined as

$$(2) \quad D(P||Q) = \sum_i p_i \log_2 \frac{p_i}{q_i}.$$

Milin *et al.* (2008a) observed, using visual lexical decision, that reaction times were longer the more a given noun's inflectional probability distribution diverged from the probability distribution of its inflectional class. For derivational mini-paradigms and classes (e.g., words with the derivational prefix *un-* and their base words), Milin *et al.* (2008b) also report effects of relative entropy.

Table 3 lists the estimated coefficients of the full regression model for Experiment 1, and Figure 5 visualizes the effects of the predictors. The random-effects part of the model contained random intercepts for lexeme ($\hat{\sigma} = 0.078$) and subject ($\hat{\sigma} = 0.070$), as well as by-subject random slopes for the linear
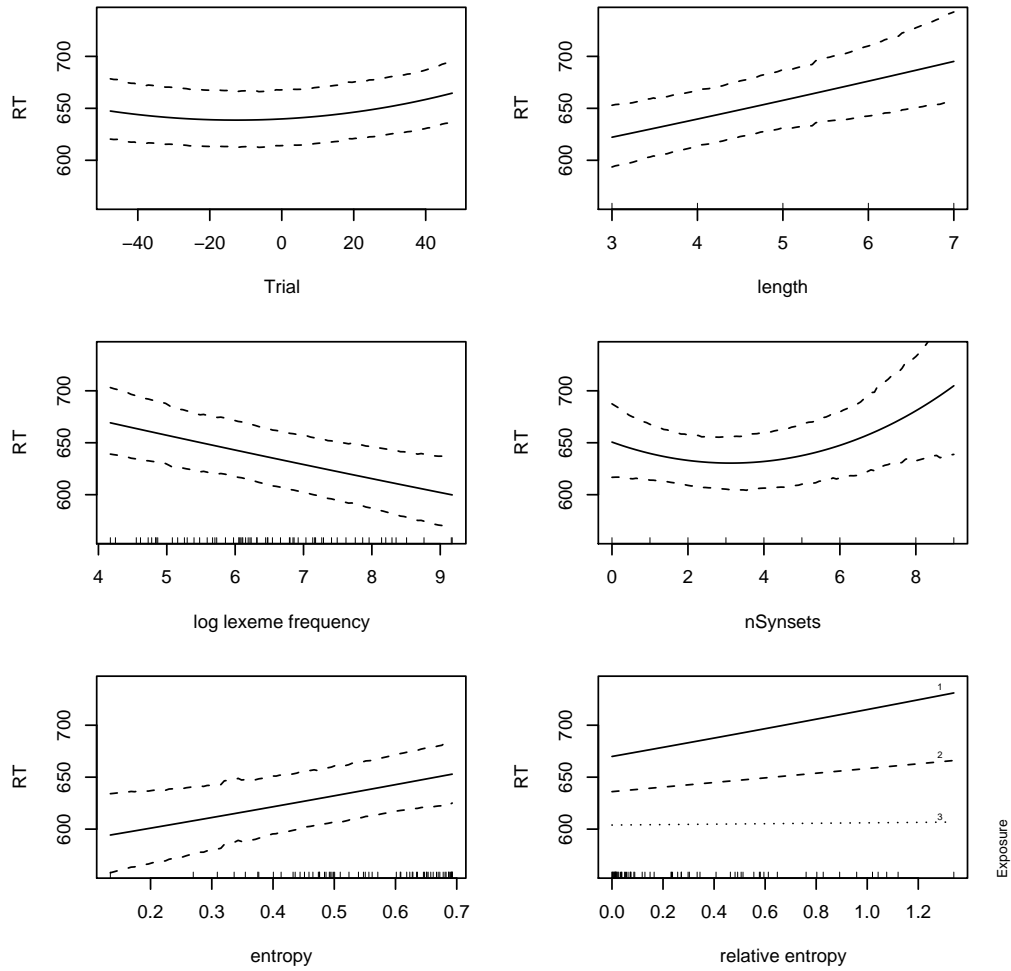
Figure 5: The partial effects of the predictors in a linear mixed-effects regression model fitted to the data of Experiment 1 (Picture Naming). Each panel is adjusted for the median value of the other predictors. Dashed lines in the upper panels and in the lower left panel denote 95% Highest Posterior Density intervals calculated from the MCMC posterior estimates of the parameters.

effect of Trial ($\hat{\sigma} = 0.070$, likelihood ratio test $p < 0.001$). The estimate for the standard deviation for the residual error was 0.115.

As expected, Length was inhibitory (upper right panel), and Lexeme Frequency was facilitatory (central left panel). The effect of the number of synsets was most prominent for higher synset counts, for which it was inhibitory. The effect of Inflectional Entropy was also inhibitory, and represents the non-dichotomized plural dominance effect. Plural dominant nouns have high-information paradigms, and apparently high-information paradigms are costly to access in speech production. The lower right panel of Figure 5 plots the effect of Relative Entropy, which was inhibitory during the first exposure, but disappeared with successive exposures. Within the context of the experiment, subjects became more familiar with the pictures and the names they were expected to produce for these pictures. Initial problems associated with retrieving a paradigm that diverges from the general class disappeared with increasing familiarity.

Before attempting to interpret these findings in further detail, we first discuss three additional experiments that seek to provide further constraints on the range of possible interpretations. Experiment 2 used delayed picture naming in order to investigate whether the effect of dominance/entropy is a late effect that arises after preparation for articulation has been completed.

3 Experiment 2: Delayed Picture Naming
3.1 Method

Materials The materials were identical to those of Experiment 1.

Subjects Thirty-two subjects, students at the university of Nijmegen, were paid to participate in the experiment. All had normal or corrected-to-normal vision and no known speech impairment. None participated in any of the other experiments reported in this study.

Procedure The procedure was identical to that of Experiment 1, except that subjects were instructed to wait for a response cue (a beep presented over headphones) before naming the picture. For the targets, the response cue was presented after 1000 ms. For one half of the fillers, the response cue was presented at 1300 ms, for the other half at 1600 ms. Naming latencies were measured from the onset of the response cue.

3.2 Results and discussion

For 4% of the data, subjects named another word or the voice key was triggered prematurely. After removal of these data points, we inspected the distribution of the naming latencies. Removal of latencies exceeding 665 ms (1.5% of the data points) and a logarithmic transformation reduced substantial skewing and resulted in a data set with an approximately normal distribution of naming latencies. A linear mixed-effects model was fitted to this data set with Subject and Lexeme as random effects. Model criticism led to removal of potentially harmful outliers (defined as data points with absolute standardized residuals exceeding 2.5), after which we refitted the model. The only

	Estimate	lower HPD	upper HPD	p (MCMC)	Pr (t)
Intercept	5.9846	5.8363	6.1159	0.0001	0.0000
Number: sg	-0.0515	-0.1004	-0.0001	0.0526	0.0034
Length	-0.0331	-0.0599	-0.0046	0.0244	0.0000

Table 4: Coefficients, highest posterior density intervals, and p-values for the fixed-effect predictors that reached significance in Experiment 2 (Delayed Naming).

	Estimate	MCMC mean	lower HPD	upper HPD
lexeme: random intercepts	0.033	0.071	0.055	0.092
subject: random intercepts	0.133	0.075	0.055	0.103
subject: Number pl	0.046	0.068	0.048	0.096
subject: Number sg	0.069	0.084	0.061	0.116
subject: correlation sg,pl	0.048	0.017 -	0.488	0.466
residual error	0.175	0.175	0.170	0.180

Table 5: Random intercepts, contrast coefficients, and correlations in Experiment 2 (Delayed Naming), with Highest Posterior Density intervals. All random intercepts and contrasts are supported by likelihood ratio tests ($p < 0.001$).

predictors that reached significance in the trimmed model were Number and Word Length. Singulars had an average processing advantage compared to plurals of 7 ms. Furthermore, longer words elicited shorter response latencies. The average latency for the longest words (7 phonemes) was 45 ms shorter than the average latency for the shortest words (3 phonemes). Coefficients and highest posterior density intervals are listed in Table 4. Table 5 lists the random-effects parameters, which include random contrasts for Number. The Markov chain Monte Carlo samples from the posterior distribution suggested that the correlation parameter for the contrasts for singular and plural was superfluous, but a likelihood ratio test argued against its removal from the model ($p < 0.0001$). Its presence or absence in the model does not affect the evaluation of the fixed-effects.

Comparing Experiments 1 and 2, we see that the effects of lexeme frequency, entropy and relative entropy, and semantic ambiguity are predictive only in immediate picture naming. In delayed picture naming, we only see a tiny effect of Number, and a facilitatory effect of word length. This inverse word length effect may arise due to the suppression of the normal processes for articulation required in delayed naming. If shorter words require stronger suppression, and if the time required to undo the suppression is proportional to the amount of suppression, then it follows that it would take more time to initiate the pronunciation of higher-frequency and shorter words, leading to the observed anti-length effect. Crucially, the absence of effects of entropy and relative entropy suggests that these two effects do not arise during articulation.

The next question to be addressed is whether the entropy effects arise during conceptualization and picture interpretation, preceding access to the lexeme. To answer this question, we used a category decision task in which subjects had to decide whether the line drawing presented on the computer

screen represented an existing object. This is a conceptual task that does not require linguistic encoding. If entropy and relative entropy effects emerge in this task, we can draw the conclusion that their effect is conceptual rather than linguistic in nature.

4   Experiment 3: Picture verification
4.1   Method

Materials We used the same pictures as in Experiment 1. To these materials, we added 32 fillers, 16 line drawings of singular objects, and 16 line drawings of plural objects. The fillers were all pictures representing non-existing figures (characters from Dutch comics such as Tom Poes and Wiske, characters from international comics (Snoopy, Obelix) and typical characters from fairy tales (giant, fairy).

Subjects Thirty-two subjects, students at the university of Nijmegen, were paid to participate in the experiment. All had normal or corrected-to-normal vision and no known speech impairment. None had participated in any of the other experiments reported in this study.

Procedure The procedure was identical to that of Experiment 1, except that we asked subjects to decide, as quickly and as accurately as possible, whether the picture presented on the computer screen represented an existing character or object by means of button presses. Button presses from the dominant hand indicated that the picture represented an existing object (e.g., 'goat'), button presses from the non-dominant hand indicated that the picture represented an imaginary object or character (e.g., 'Snoopy'). Response latencies were measured from the onset of the picture.

4.2   Results and discussion

Incorrect responses and responses exceeding the timeout of 2000 ms were classified as errors. There were very few errors, 49 on a total of 3072 (1.6%). For the analysis of the decision latencies, we inspected the order statistics of the distribution of decision latencies, and removed 54 extreme data points (1.8% of the correct responses) for which the latency exceeded 1000 ms. After removal of these outliers, the distribution of log-transformed decision latencies was approximately normal. We fitted a linear mixed-effects regression model to the data with lexeme and subject as random effects, using a stepwise variable elimination procedure. Potentially overly influential outliers (defined as observations with absolute standardized residuals exceeding 2.5) were removed from the data set, after which the model was refitted. Table 6 lists the coefficients of the fixed effect predictors together with their Highest Posterior Density intervals, Table 7 lists the standard deviations estimated for the random effects, which included random slopes for Exposure ($p < 0.001$, likelihood ratio test).

Lexeme Frequency was predictive for the category decision latencies, and facilitatory. Subjects responded more quickly with each successive exposure. A small ambiguity effect, as gauged by the synset count, was attenuated with

	Estimate	lower HPD	upper HPD	p (MCMC)	Pr (t)
(Intercept)	6.4426	6.3194	6.5723	0.0001	0.0000
Frequency	-0.0178	-0.0362	0.0007	0.0562	0.0137
Exposure	-0.0481	-0.0722	-0.0237	0.0006	0.0000
Synset Count	0.0108	-0.0046	0.0256	0.1738	0.0833
Exposure by Synset Count	-0.0047	-0.0079	-0.0013	0.0092	0.0069

Table 6: Coefficients, highest posterior density intervals, and *p*-values for the fixed-effect predictors that reached significance in Experiment 3 (Category Verification).

	Estimate	MCMC mean	lower HPD	upper HPD
by-lexeme random intercepts	0.0657	0.08715	0.07163	0.1055
by-subject random intercepts	0.1391	0.10387	0.07858	0.1376
by-subject random slopes exposure	0.0273	0.06191	0.04535	0.0872
residual error	0.1263	0.12585	0.12277	0.1294

Table 7: Random intercepts, contrast coefficients, and correlations in Experiment 3 (Category Verification), with Highest Posterior Density intervals.

each exposure. The entropy measures did not reach significance. The presence of a lexeme frequency effect and a synonym effect, combined with the absence of entropy effects, suggests that the entropy effects arise after picture interpretation and conceptualization. This combination of effects also suggests that the lexeme frequency effect taps, at least in part, into conceptual familiarity.

Experiments 1–3 made use of simple line drawings. For the elicitation of plurals, we made use of two pictures placed side by side. For nouns such as *mond* ('mouth'), the resulting pictures were odd. Experiment 4 was designed to avoid such unnatural stimuli by making use of photographs of objects. In addition to single objects, pictures of two objects as well as pictures of more than two objects were included, in order to evaluate whether the results of Experiment 1 might have been influenced by only two objects having been shown for the plural. Another change with respect to the first three experiments was that subjects were not familiarized with the intended picture names before the experiment, in order to avoid possible interference from prior familiarization, which might prime the relevant representations in memory.

5   Experiment 4: Photograph Naming
5.1   Method

Materials For each of 101 picturable nouns, six photographs were made following the design shown in Figure 6.[2] Three photographs showed the target object against a natural background (left column), and three photographs showed the object against a neutral background. Removal of the background often required repositioning the objects, as illustrated here for plates. In what follows, the factorial contrast between the presence versus absence of a natural background is referred to as Context. For each context, photographs were made of one object (top row), two objects (center row) and more than two

---

[2]We are indebted to Laurens Krol for the creation of this photograph database.

objects (bottom row). In what follows, we refer to these levels of the factor Number as singular, dual, and plural. The 606 pictures were randomly assigned to six lists following a 6 by 6 Latin Square design. Each list contained 101 pictures, with approximately equal counts (16 or 17) of each combination of Number and Context.

Figure 6: An example of the kind of photographs used in Experiment 4 (Photograph Naming): singulars (top), duals (center) and plurals (bottom) with (left) and without (right) context.

For each noun we calculated, using the CELEX lexical database, the log lexeme frequency, the log frequency of the singular form, the log of the frequency of the plural form, the log of the morphological family size, the log of the number of synsets in WordNet. We also calculated the inflectional entropy of the word's number paradigm and its relative entropy, using as reference distribution the probabilities of the singular ($p = 0.75$) and plural ($p = 0.25$) estimated from monomorphemic nouns in CELEX. Various other variables pertaining to word length and neighborhood density were also considered, but as they did not reach significance in the statistical analysis, we do not discuss them any further. (Neighborhood density did not reach significance in the

preceding experiments either.)

Subjects Sixty students at the university of Nijmegen were paid to participate in this experiment. All had normal or corrected to normal vision, and no known speech impairment. Each of the six lists of the Latin Square was assigned ten subjects.

Procedure Subjects were tested one by one in a noise-attenuated experimentation booth. A fixation mark was presented in the center of the screen for 1000 ms. After 50 ms, the photograph was shown, in portrait mode, using the full vertical dimension of the computer screen. Photographs remained on the screen for 3000 ms. A new trial was initiated 500 ms afterwards. There were six short breaks during the experiment, one after each block of 101 photographs. The total duration of an experimental session was approximately 60 minutes.

5.2 Results and discussion

Due to technical failures of the experimental software (which was taxed to its limits by the large number of bitmaps for the photographs that had to be loaded in memory), the data of four subjects were lost. Inspection of the distribution of naming latencies for the remaining 56 subjects showed a marked departure from normality. We reduced the skew by removing 232 trials with naming latencies less than 400 ms and 1466 trials with latencies exceeding 1500 ms (5% of the data points), followed by a logarithmic transform.

	Estimate	lower HPD	upper HPD	p (MCMC)	p (t)
Intercept	6.9085	6.8460	6.9697	0.0001	0.0000
Trial (linear)	0.0019	0.0016	0.0022	0.0001	0.0000
Trial (quadratic)	0.0000	0.0000	0.0000	0.0001	0.0000
Context present	0.0223	0.0144	0.0304	0.0001	0.0000
Lexeme Frequency	-0.0317	-0.0385	-0.0255	0.0001	0.0000
Relative entropy	-0.0291	-0.0609	-0.0013	0.0460	0.3399
Entropy	0.0378	-0.0088	0.0909	0.1160	0.4804
Plural TRUE	0.0727	0.0373	0.1088	0.0001	0.0000
Dual TRUE	0.0286	0.0150	0.0416	0.0002	0.0000
Block	-0.0249	-0.0338	-0.0145	0.0001	0.0000
Relative Entropy: plural	-0.0369	-0.0607	-0.0128	0.0018	0.0000
Entropy: plural	-0.0487	-0.0904	-0.0064	0.0238	0.0020
Blocks: dual	-0.0052	-0.0078	-0.0028	0.0001	0.0000

Table 8: Coefficients of the mixed-effects regression model fitted to the data of Experiment 4 (Photograph Naming). The reference level for Context is 'absent', for Dual and Plural the reference level is 'FALSE'.

We fitted a linear mixed-effects regression model to the data with log naming latency as the dependent variable, with Trial and Block as control predictors, and with Context, log Lexeme Frequency, Number, Entropy, Relative Entropy, and Number of Synsets as predictors of interest. Instead of including Number as a three-level factor, we included separate predictors for whether the picture represented two objects (Dual, levels TRUE/FALSE) or more than
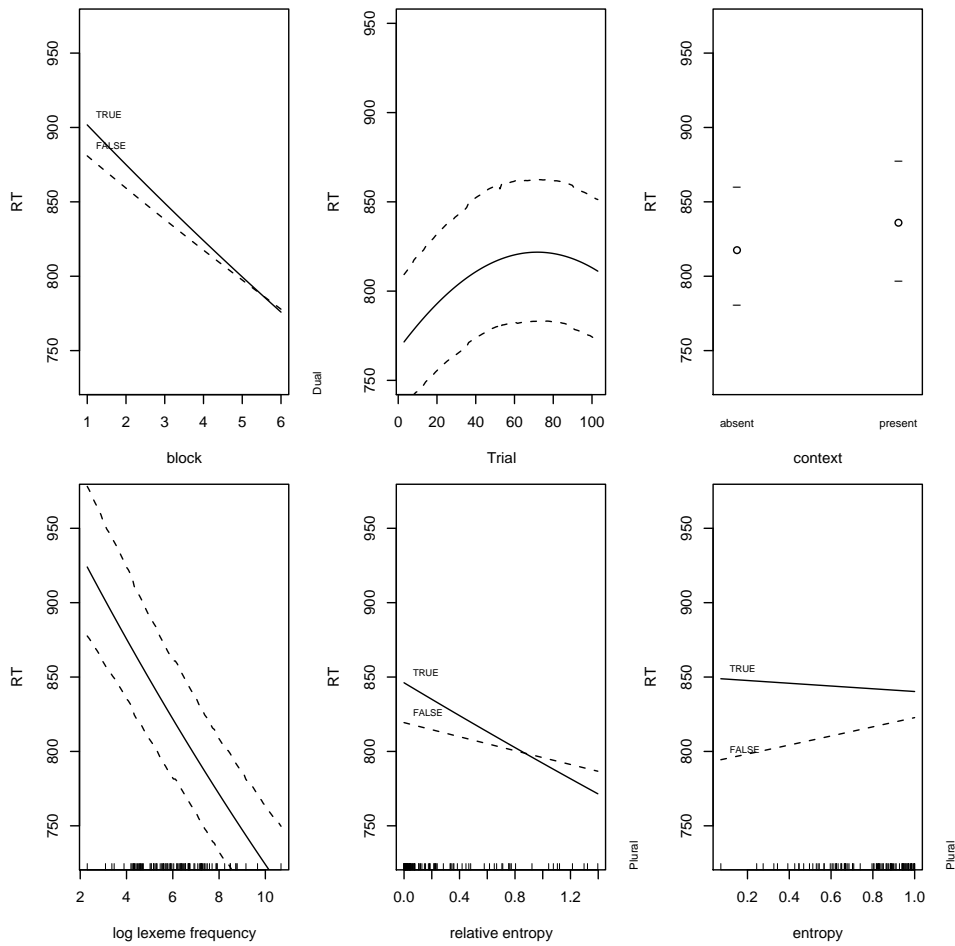
Figure 7: The partial effects of the fixed-effects predictors in a linear mixed-effects model fitted to the data of Experiment 4 (Photograph Naming). Each graph is adjusted for the median value of the other covariates. Dashed lines represent 95% Highest Posterior Density intervals (for panels without interactions only).

two objects (Plural, TRUE/FALSE), as this facilitated parsimoneous modeling of interactions involving number.

Random-effect factors were Photograph, Lexeme, and Subject. Table 8 summarizes the estimated coefficients of the fixed effects predictors that reached significance, and Figure 7 visualizes the corresponding partial effects. Finally, Table 9 lists the random effect parameters and associated statistics. In addition to random intercepts for Picture, Lexeme, and Subject, the model included by-word random slopes for Blocks, by-subject random slopes for Blocks, and by-subject random slopes for Relative Entropy. All random parameters were supported by likelihood ratio tests (all $p < 0.0001$). The by-word and by-subject random slopes for Blocks bring processing differences that probably arise due to the use of a Latin Square design into the model: a given block provides the context for any of the photographs in that block. The presence of by-subject random slopes for Relative Entropy in the model shows that for this predictor there is significant variability among subjects. We return to this variability below.

	Estimate	lower HPD	upper HPD
photograph: random intercepts	0.04448	0.04068	0.04822
lexeme: random slopes Blocks	0.04435	0.03716	0.05287
lexeme: random intercepts	0.04127	0.03397	0.04990
subject: random slopes RH	0.02523	0.01813	0.03456
subject: random slopes Blocks	0.01485	0.01211	0.01833
subject: random intercepts	0.10207	0.08431	0.12426
residual error	0.17328	0.17185	0.17472

Table 9: The estimated random effects parameters (standard deviations) in the mixed-effects model fitted to the naming latencies of Experiment 4 (Photograph Naming), as well as 95% Highest Posterior Density intervals, based on 10000 Markov chain Monte Carlo samples from the posterior distribution of the parameters. RH: relative entropy.

Of the fixed-effects factors, we first discuss the two control variables. The mean naming latency decreased with each successive block, from some 900 ms in the first block to some 800 ms in the last block (upper left panel). This effect was slightly stronger for the duals (represented by the solid line). As subjects went through a given block, their naming latencies became longer, reaching a plateau after some 70 trials into the block (upper central panel). Recall that each new block was preceded by a break. Apparently, subjects started each new block with new vigor, but then grew tired as they progressed through the block and named the pictures presented to them more slowly.

The upper right panel shows the effect of Context. Naming latencies were slightly shorter for the objects presented without context. The disadvantage for objects presented in context is probably due to the greater interpretational complexity of these pictures. (We note here that the complexity of the individual pictures, as gauged by the file sizes of their jpg files, was not predictive.

The random effect of Photograph in the model probably fully captures the relevant differences in picture complexity.) Context did not interact with any of the other predictors. This suggests that it is not necessary to carry out the labor-intensive image manipulation required to decontextualize objects in photographs.

From the coefficients for Plural and Dual available in Table 9 it is clear that plurals elicited longer naming latencies ($\hat{\beta} = 0.0727$) than singulars. Duals also elicited somewhat longer latencies than singulars ($\hat{\beta} = 0.0286$, but this difference disappeared in later blocks (see the upper left panel of Figure 7).

The first panel on the second row of Figure 7 shows the effect of log lexeme frequency. The more frequent the name of an object is, the shorter the latency to its picture. This lexeme frequency effect may reflect conceptual familiarity (compare Experiment 3), familiarity with the lexeme, or both.

The final panels on the bottom row of Figure 7 visualize the effects of Relative Entropy and Entropy. The effect of Relative Entropy is present only for the plurals. The effect for the non-plurals has only weak support: the p-value based on the t-distribution is not signficant, and the p-value based on the posterior distribution of the parameter is just significant at the 5% level. Furthermore, we should take into account that, as mentioned above, the model incorporates random slopes for Relative Entropy. Working from the model coefficients, and the range of the BLUPs for the by-subject random slopes (-0.041 to 0.053), we find that the slopes of individual subjects range from -0.119 to -0.024 for plurals and from -0.070 to 0.024 for singulars. In short, it is only for the plurals that Relative Entropy was consistently facilitatory.

The effect of Entropy also varies between plurals and non-plurals. The bottom right panel of Figure 7 illustrates this interaction. For the non-plurals, its effect is inhibitory. For the plurals, there is not much of an effect at all. In summary, for plurals we have facilitation from Relative Entropy, for non-plurals, we have inhibition from Entropy.

The inhibitory effect of Inflectional Entropy replicates the inhibitory effect of Entropy observed for Experiment 1. Recall that Experiment 1 presented only singular and dual pictures, exactly the kind of pictures for which we observe inhibition in Experiment 4. The pattern of results for Relative Entropy is less clear. In Experiment 1, an initial inhibitory effect for the first exposure disappeared by the third exposure. In Experiment 4, there was no evidence for an interaction of Relative Entropy by Block. Across six blocks, Relative Entropy was not predictive for the subset of singulars and duals, mirroring its non-significance for singular and dual stimuli at the third exposure in Experiment 1.

The present findings raise two questions. First, why is it that duals side with singulars and not with plurals? Second, why is Relative Entropy predictive for only the plurals, and Entropy only for the singulars and duals?

For singulars and duals, Entropy emerges as the relevant predictor. Here the question is why Entropy is not predictive when pictures with plural objects have to be named. A tentative answer proceeds by noting that there are two differences between the pictures of plurals and the pictures for the non-plurals.

First, plural pictures always present more than two objects. Second, our plural pictures present a varying number of objects. In other words, the dual and the singular pictures in Experiment 4 have in common that they display a specific number of objects (one or two) whereas the plural pictures reference an unspecified, varying number of objects. It might be that the effect of Inflectional Entropy is restricted to the singulars and duals because here the speaker is confronted with a specific number of objects (*one shoe, two shoes; one marble, two marbles*) rather than with more objects (*several pairs of shoes; a handful of marbles*). Possibly, it is this numerical specificity that motivates, at least in part, the use of a special dual form (for English, compare the series *the shoe, both shoes, several shoes*, where *both* is a dual form).

This still leaves us with the question why relative entropy is predictive primarily for the plurals. Recall that relative entropy quantifies the extent to which the probability distribution of a given lexeme diverges from the generalized probability distribution of the class of nouns. For plurals, a greater relative entropy indicates a greater divergence from the 'norm' of being singular dominant. In other words, for plurals, a greater relative entropy is an index of a high degree of plural dominance. The plurals in Experiment 2 emerged with a negative slope for Relative Entropy. This indicates that, other things being equal, greater plural dominance affords shorter picture naming latencies for pictures with more than two objects. Possibly, the less a lexeme's P distribution approximates the general Q reference distribution, the less its plural form is in the gravitational field of the singular, and the faster the plural can be articulated.

6   Comprehension of number revisited

Given the results for speech production and the importance of paradigmatic structure for the processing of inflected forms, and given the methodological problems associated with dichotomization of numeric variables, we decided to revisit the comprehension of singular and plural forms as reported by Baayen *et al.* (1997b) (Experiment 1). Above, we summarized these visual lexical decision data visually (Figure 1). The upper part of Table 10 lists the six coefficients required in the factorial mixed-effects model fitted to these data (with subject and lexeme as random effects). The lower part of this table lists the four coefficients in a mixed-effects analysis of covariance fitted to exactly the same data. The estimated standard deviation of the residual error is the same for both models up to three decimal digits (0.186). The log likelihood of the more parsimoneous model (1923) with four fixed-effects coefficients is greater than that of the model with six fixed effects coefficients (1913). An interaction plot based on the fitted values of the parsimoneous model is indistinguishable from the interaction plot shown in Figure 1. This allows us to conclude that the analysis of covariance is superior: it provides a simpler model without loss of accuracy.

In this more parsimoneous model, the factor Lexeme Frequency is replaced by the corresponding by-item lexeme frequencies. Number, a genuine factor,

	original factorial analysis				
	Estimate	lower HPD	upper HPD	p (MCMC)	p (t)
Intercept	6.3007	6.2630	6.3383	0.0001	0.0000
Number sg	-0.0220	-0.0357	-0.0093	0.0016	0.0012
Dominance sgdom	0.0834	0.0510	0.1158	0.0001	0.0000
Frequency low	0.1320	0.1000	0.1650	0.0001	0.0000
Number sg : Dom sgdom	-0.0683	-0.0841	-0.0528	0.0001	0.0000
Number sg : Freq low	-0.0405	-0.0561	-0.0249	0.0001	0.0000
	reanalysis with analysis of covariance				
	Estimate	lower HPD	upper HPD	p (MCMC)	p (t)
Intercept	6.6373	6.5698	6.7117	0.0001	0
Information	0.0273	0.0220	0.0323	0.0001	0
Number sg	-0.0530	-0.0616	-0.0440	0.0001	0
Frequency	-0.0407	-0.0508	-0.0312	0.0001	0

Table 10: Coefficients for mixed-effects reanalyses of Experiment 1 of Baayen et al. (1997) (visual lexical decision). The upper half replicates the original factorial analysis. The lower half presents a re-analysis using analysis of covariance. For both models, the estimate of the standard deviation of the residual error is 0.186. The log likelihood is 1913 for the factorial model and 1923 for the reanalysis with analysis of covariance.

is retained. The third predictor in this model is the information carried by the inflected form in its paradigm. Given the frequencies F_{sg} and F_{pl} for lexeme L, the probabilities of the singular and the plural given L are

$$
\begin{aligned}
\Pr(sg|L) &= \frac{F_{sg}}{F_{sg} + F_{pl}} \\
(3) \qquad \Pr(pl|L) &= \frac{F_{pl}}{F_{sg} + F_{pl}}
\end{aligned}
$$

and the corresponding amounts of information are

$$
\begin{aligned}
I_{sg|L} &= -\log_2(\Pr(sg|L)) \\
(4) \qquad I_{pl|L} &= -\log_2(\Pr(pl|L)).
\end{aligned}
$$

These information estimates are, in fact, a simplification of the measure proposed by Kostić *et al.* (2003) for the analysis of the processing costs of nominal case in Serbian, the main difference being that we did not weight for the number of functions and meanings of an inflected form. Note that the coefficient of an inflected form's amount of information listed in Table 10, 0.0273, is positive. The greater the amount of information that has to be retrieved from the mental lexicon, the longer its response latencies will be. Equivalently, the greater the relative frequency of a form in its number paradigm, the shorter its response latencies will be. (It is noteworthy that, unlike singular or plural frequency, the amount of information in not significantly correlated with lexeme frequency ($r < 0.1$).) It is not the absolute frequency of the inflected form by itself that is at issue, but its relative frequency in the number paradigm. The greater its relative frequency, the more accessible an inflected form is in its paradigm.

To complete the reanalysis of the lexical decision data, we included Entropy, Relative Entropy and Family Size as predictors, and also considered potential nonlinearities. The resulting model is summarized in Table 11 and visualized in Figure 8. Upon closer inspection, the effect of lexeme frequency turned out to be nonlinear, leveling off for the higher frequencies. Lexeme frequency also was somewhat less facilitatory for singulars compared to plurals. There was some support for facilitatory effects of Family Size and Relative Entropy. There was no evidence supporting an effect of Entropy.

	Estimate	lower HPD	upper HPD	p (MCMC)	p (t)
Intercept	6.9621	6.6986	7.2196	0.0001	0.0000
Information	0.0243	0.0197	0.0289	0.0001	0.0000
Number sg	-0.1392	-0.1703	-0.1080	0.0001	0.0000
Family Size	-0.0167	-0.0349	0.0008	0.0624	0.0205
Lexeme Freq (linear)	-0.1279	-0.2022	-0.0506	0.0018	0.0000
Lexeme Freq (quadr.)	0.0066	0.0015	0.0117	0.0124	0.0015
Relative Entropy	-0.0533	-0.1063	0.0038	0.0654	0.0176
Number sg by Freq	0.0126	0.0079	0.0171	0.0001	0.0000

Table 11: Coefficients for mixed-effects reanalyses of Experiment 1 of Baayen et al. (1997) (visual lexical decision), using a broader range of predictors.

When we consider the production and comprehension data jointly, an important similarity is the presence of a facilitatory lexeme frequency effect. This suggests that in both modalities an abstract representation for the lexeme (in the sense of Aronoff (1994), or a lemma in the sense of Levelt (1989)) is accessed first. What is different between the two modalities is the subsequent access to the inflected forms given the lexeme. In comprehension, it is the probability of the singular (or plural) form given the lexeme that comes into play. The greater an inflected form's paradigmatic probability, the faster it is accessed. In production, it is the entropy of the paradigm that is at issue, at least for singulars and duals. The inhibitory effect of entropy may reflect the cost of accessing or activating the paradigm.

7   Discussion

The main goal of the present study was to trace the consequences of local and global markedness for the processing of singular and plural nouns. Decompositional models such as proposed by (Pinker (1997); Pinker (1999)) and (Levelt *et al.* (1999)) predict a lexeme frequency effect and no effects of the frequencies of the singular and the plural forms. Experiments 1 and 4 revealed the expected lexeme frequency effect. Furthermore, in these experiments there were no clear independent effects of the frequencies of the inflected forms. However, the effects of Entropy and Relative Entropy that emerged from these experiments show that in production knowledge of the probabilities of the individual inflected forms do play a role, albeit indirectly. These entropy effects bear witness to the importance of paradigmatic organization of inflected forms in the mental lexicon, both at the level of individual lexemes (Entropy) and at the general level of the class of nouns (Relative Entropy).
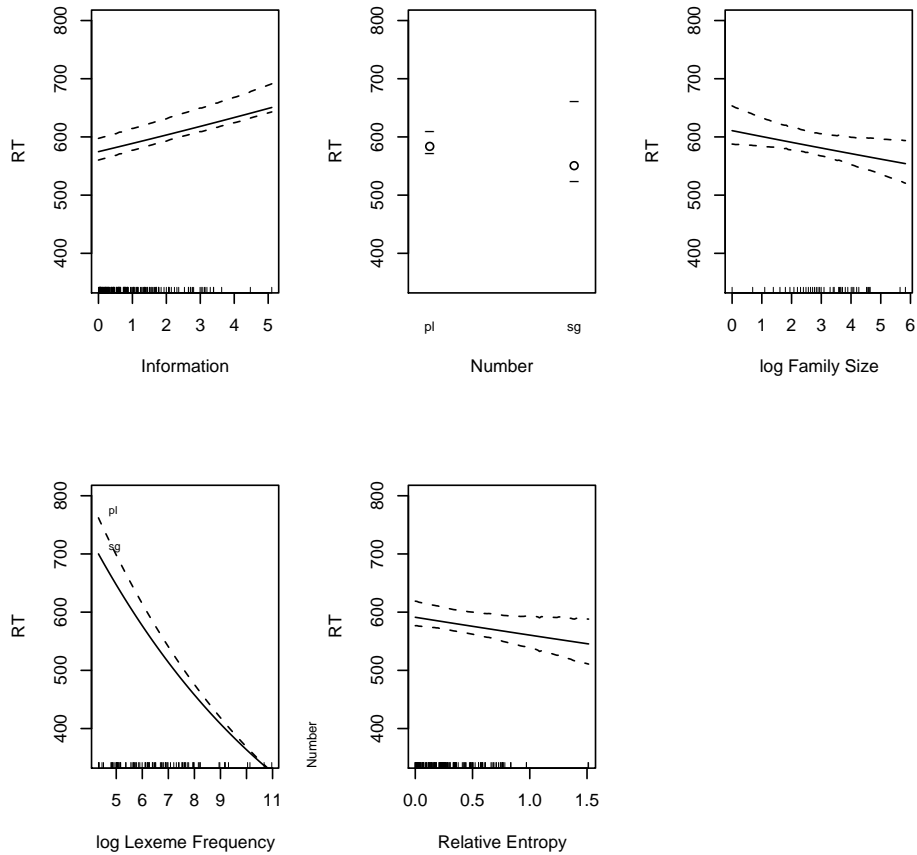
Figure 8: Partial effects of the predictors in the full model for the visual lexical decision latencies in Experiment 1 of Baayen, Dijkstra & Schreuder (1997). 95% Highest Posterior Density credible intervals are shown for all panels except Lexeme Frequency, for which the interaction with Number is graphed.

The present results are compatible with Word and Paradigm morphology (see, e.g., Matthews (1974); Blevins (2003); Blevins (2006)). Word and Paradigm morphology can be viewed as an exemplar theory of (inflectional) morphology, in which rules are analogical generalizations across exemplars in lexical memory. The effect of Entropy presupposes knowledge of the likelihood of inflectional exemplars, the effect of Relative Entropy supports the hypothesis that there is generalized knowledge (the Q distribution) against which lexeme-specific information (the P distribution) is evaluated. In other words, the effect of Relative Entropy may reflect the processing costs of analogical generalization.

The present data suggest that processing models that account for regular inflection only through rule-based derivation from the stem underestimate the complexity of the organization of the mental lexicon. Just as linguistic theories have often ignored the importance of paradigmatic relations, (de)compositional processing theories have never considered seriously the possibility of a hierarchically structured organization of lexemes and their paradigms. In a hierarchically structured lexicon, the Lexeme Frequency effect can be seen as reflecting the costs of accessing a given lexeme, the Entropy effect as reflecting the cost of accessing the lexeme's paradigm, and the Relative Entropy effect as reflecting the cost of applying a general analogical rule. In short, as for comprehension, speech production requires a model in which storage and computation work in synergy (cf. Levelt (1989)).

A subsidiary goal of the present paper was to show that concepts from information theory may contribute substantially to our understanding of lexical representation and processing. The effects of Entropy and Relative Entropy that we observed in Experiments 1 and 4 show that information theory indeed provides good conceptual tools for predicting processing costs in speech production. The reanalysis of the data of Baayen *et al.* (1997b) suggested furthermore that Relative Entropy may also play a role in language comprehension.

A final methodological goal of this paper was to illustrate that dichotomization of numeric predictors is not helpful. The original dichotomization of dominance in Experiment 1 led us to hypothesize that plural-dominant nouns would have two entries in the lemma stratum of the WEAVER model of speech production (Levelt *et al.* (1999)) whereas singular-dominant nouns would have only one entry. The hypothesis of Levelt et al. (1999) is that (singular-dominant) nouns with only one entry would allow faster lexical access than (plural-dominant) nouns with two (nearly synonymous) entries, for which a selection problem would have to be resolved. However, since dominance is a graded notion, this approach raises the question how plural-dominant a noun would have to be for it to be assigned a second lemma. This problem of discretization is no longer an issue once Dominance is replaced by Inflectional Entropy. Furthermore, whereas Levelt *et al.* (1999) explained the effect of Dominance as a choice problem between singular and plural form at the stratum of lemma representations, the present data and analyses suggest that this choice problem probably arises at a lower level, within a lexeme's morphological paradigm.

For language comprehension, we have seen that a much simpler explanation for the results of Baayen *et al.* (1997b) is obtained by replacing the factor Dominance by the relative frequency of an inflected form in its paradigm, and by replacing the factor Lexeme Frequency by the actual lexeme frequencies of the lexemes used in the experiment. A reanalysis along similar lines of the data reported by New *et al.* (2004) for English and French may help resolve the discrepancy between the effect of Dominance they observed for these two languages.

References

Aronoff, M. 1994. *Morphology by itself: stems and inflectional classes*. Cambridge, Mass.: The MIT Press.

Baayen, R. H. 2008. *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

——, C. Burani, & R. Schreuder. 1997a. Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian. In *Yearbook of Morphology 1996*, ed. by G. E. Booij & J. van Marle, 13–34. Dordrecht: Kluwer Academic Publishers.

——, D. J. Davidson, & D.M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, in press.

——, T. Dijkstra, & R. Schreuder. 1997b. Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language* 36.94–117.

——, L. Feldman, & R. Schreuder. 2006. Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language* 53.496–512.

——, J. McQueen, T. Dijkstra, & R. Schreuder. 2003. Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In *Morphological structure in language processing*, ed. by R. H. Baayen & R. Schreuder, 355–390. Berlin: Mouton de Gruyter.

——, R. Schreuder, N. H. De Jong, & A. Krott. 2002. Dutch inflection: the rules that prove the exception. In *Storage and Computation in the Language Faculty*, ed. by S. Nooteboom, F. Weerman, & F. Wijnen, 61–92. Dordrecht: Kluwer Academic Publishers.

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

Bates, D. M. 2005. Fitting linear mixed models in R. *R News* 5.27–30.

Blevins, J. P. 2006. English inflection and derivation. In *Handbook of English Linguistics*, ed. by B. Aarts & A. M. McMahon, 507–536. London: Blackwell.

Blevins, James P. 2003. Stems and paradigms. *Language* 79.737–767.

Cohen, J. 1983. The cost of dichotomization. *Applied Psychological Measurement* 7.249–254.

Dimmendaal, G. J. 1987. Drift and selective mechanisms in morphological changes: the Eastern Nilotic case. In *Papers from the 7th international conference on*

historical linguistics, ed. by A. G. Ramat, O. Carruba, & G. Bernini, 193–210. Amsterdam: John Benjamins.

Ernestus, M., & R. H. Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79.5–38.

Fellbaum, C. (ed.) 1998. *WordNet: An electronic database*. Cambridge, MA: The MIT Press.

Jastrzembski, J. 1981. Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology* 13.278–305.

Jescheniak, Jörg D., & Willem J. M. Levelt. 1994. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition* 20.824–843.

Keuleers, E., D. Sandra, W. Daelemans, S. Gillis, G. Durieux, & E. Martens. 2007. Dutch plural inflection: The exception that proves the analogy. *Cognitive Psychology* 54.283–318.

Kostić, Aleksandar, Tanja Marković, & Aleksandar Baucal. 2003. Inflectional morphology and word meaning: Orthogonal or co-implicative domains? In *Morphological Structure in Language Processing*, ed. by R. Harald Baayen & Robert Schreuder, 1–44. Berlin: Mouton de Gruyter.

Kuperman, V., R. Bertram, & R. H. Baayen. 2008. Morphological dynamics in compound processing. *Language and Cognitive Processes*, to appear.

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

——, A. Roelofs, & A. S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.1–38.

MacCallum, R.C., S. Zhang, K.J. Preacher, & D.D. Rucker. 2002. On the practice of dichotomization of quantitative variables. *Psychological Methods* 7.19–40.

Matthews, P. H. 1974. *Morphology. An Introduction to the Theory of Word Structure*. London: Cambridge University Press.

Maxwell, S. E., & H. D. Delaney. 1993. Bivariate median splits and spurious statistical significance. *Psychological Bulletin* 113.181–190.

Milin, P., D. Filipović Đurđević, & Fermin Moscoso del Prado Martín. 2008a. The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from serbian. *Journal of Memory and Language*, in press.

——, V. Kuperman, A. Kostić, & R.H. Baayen. 2008b. Words and paradigms bit by bit: An information-theoretic approach to the processing of inflection and derivation. In *Analogy in Grammar: Form and Acquisition*, ed. by J. P. Blevins & J. Blevins. Oxford: Oxford University Press, in press.

Miller, G. A. 1990. Wordnet: An on-line lexical database. *International Journal of Lexicography* 3.235–312.

Moscoso del Prado Martín, Fermín, Aleksandar Kostić, & R. Harald Baayen. 2004. Putting the bits together: An information theoretical perspective on morphological processing. *Cognition* 94.1–18.

New, B., M. Brysbaert, F. L. Segui, & K. Rastle. 2004. The processing of singular and plural nouns in French and English. *Journal of Memory and Language* 51.568–585.

Pinker, S. 1997. Words and rules in the human brain. *Nature* 387.547–548.

Pinker, Steven. 1999. *Words and Rules: The Ingredients of Language*. London: Weidenfeld and Nicolson.

Sereno, J., & A. Jongman. 1997. Processing of English inflectional morphology.

Memory and Cognition 25.425–437.

Shannon, C. E., & W. Weaver. 1949. *The Mathematical Theory of Communication.* Urbana: The University of Illinois Press.

Stemberger, J. P., & B. MacWhinney. 1986. Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition* 14.17–26.

——, & ——. 1988. Are lexical forms stored in the lexicon? In *Theoretical Morphology: Approaches in Modern Linguistics*, ed. by M. Hammond & M. Noonan, 101–116. London: Academic Press.

Tiersma, P. M. 1982. Local and General Markedness. *Language* 58.832–849.

Appendix A: Nouns studied in Experiment 1

Low lexeme frequency, singular dominant: koor (choir), peer (pear), kurk (cork), bijl (ax), harp (harp), vork (fork), kluis (safe), sfinx (sphinx), zeis (scythe), pauw (peacock), koets (coach) muts (cap), schort (apron), galg (gallows), krans (wreath), taart (pie).

Low lexeme frequency, plural dominant: piek (peak), hoef (hoof), wiek (vane), braam (bramble), mier (ant), geit (goat), nier (kidney), tulp (tulip), rups (catterpillar), gans (goose), pruim (plum), wesp (wasp), klauw (claw), wilg (willow), meeuw (gull), worm (wurm).

High lexeme frequency, singular dominant broek (trousers), helm (helmet), lamp (lamp), trein (train), pijl (arrow), kruis (cross), stoel (chair), riem (belt), krant (newspaper), jurk (dress), kraan (tap), zwaard (sword), eend (duck), troon (throne), mond (mouth), poort (gate).

High lexeme frequency, plural dominant: boon (bean), boer (farmer), laars (boot), berg (mountain), struik (bush), wolk (cloud), wiel (wheel), veer (feather), kaars (candle), voet (foot), klomp (wooden shoe), plant (plant), muis (mouse), plank (plank), schoen (shoe), bloem (flower).

Appendix B: Nouns studied in Experiment 4

aardappel (potato), aardbei (strawberry), ananas (pineapple), appel (appel), asperge (asparagus), augurk (pickle), baksteen (brick), bal (ball), banaan (banana), batterij (battery), beha (bra), beitel (chisel), beker (mug), bidon (water bottle), blaadje (leaf), bloem (flower), boek (book), boor (drill), bord (plate), borstel (brush), briefkaart (postcard), bril (spectacles), broek (trousers), cheque (cheque), citroen (lemon), dadel (date), diskette (disk), doos (box), druif (grape), duif (pigeon), ei (egg), erwt (pea), gewicht (weight), glas (glass), hamer (hammer), hand (hand), handschoen (glove), hark (rake), horloge (watch), joker (joker), kaars (candle), kam (comb), kiwi (kiwi fruit), knikker (marble), komkommer (cucumber), kraal (bead), krant (newspaper), kruidnagel (clove), kurketrekker (corkscrew), kwast (brush), laars (boot), lepel (spoon), lucifer (match), mes (knife), muis (mouse), munt (coin), muts (hat), naald (needle),

overhemd (shirt), paardebloem (dandelion), pan (pan), paperclip (paperclip), peer (pear), pen (pen), penseel (brush), pet (cap), pinda (peanut), pleister (plaster), postzegel (stam), potlood (pencil), radijs (radish), ring (ring), rok (skirt), roos (rose), schilderij (painting), schoen (shoe), schroef (screw), schroevedraaier (screwdriver), sinaasappel (orange), sleutel (key), sok (sock), steen (stone), stekker (plug), stoel (chair), stopcontact (socket), stropdas (tie), tandenborstel (toothbrush), tomaat (tomato), trui (sweater), ui (union), veer (feather), vergiet (sieve), vijl (file), vinger (finger), voet (foot), vork (fork), vuist (fist), walnoot (walnut), worst (sausage), wortel (carrot), zaag (saw).