

The Processing of Singular and Plural Nouns in English, French, and Dutch: New Insights From Megastudies

Manuel Gimenes
University of Poitiers, CNRS

Marc Brysbaert
Ghent University

Boris New
CNRS, University Savoie Mont Blanc

In this study, we explored the processing of singular and plural word forms, using megastudies in French, English, and Dutch. For singulars, we observed a base frequency effect but no surface frequency effect. For plurals, the effect depended on the frequency of the word form. When the word form had a frequency above a threshold value, we observed both surface and base frequency effects; for the frequencies below the threshold, we found a base frequency effect only, suggesting full decomposition for these words. The threshold differed between the languages, suggesting that more plurals are decomposed in French than in Dutch and more in Dutch than in English. In contrast, for all languages the singular form seems to be coactivated whenever the plural form is processed. These results are interpreted in light of some of the main models of morphological processing.

Keywords: visual word recognition, morphology, plural, megastudy

Understanding visual word recognition is an important topic in psycholinguistics. While there is general consensus that monomorphemic words are stored in the mental lexicon, there is an ongoing debate about the way in which morphologically complex words are processed. There are three main classes of models: full listing models (Butterworth, 1983), full parsing models (Clahsen, 1999; Taft, 2004), and dual-route models (Schreuder & Baayen, 1995).

Full listing models (Butterworth, 1983; Manelis & Tharp, 1977) posit that all words are stored in the mental lexicon, even morphologically complex words. According to these models, there is an independent lexical representation for each word and this implies that each inflected or derived word is represented globally and independently without any morphemic decomposition.

Full parsing models state that all morphologically complex words are decomposed. An interesting model of full parsing is the interactive-activation model of Taft (1994, 2004). The main idea is that polymorphemic words are always decomposed into individual

morphemes and are recognised via the representation of their stem. More precisely, this model postulates two stages: an early and obligatory stage of morpheme decomposition, and a late processing stage in which the functional information associated with the stem representation is recombined with the functional information associated with the separately accessed suffix representation. It can be noted that recombination on the basis of functional information occurs for inflected words but not for derived words. An important feature of the late stage of recombination is that its duration is not constant: The decision that the affix can be combined with the stem takes more or less time depending on how easy it is to establish that they go together.

Dual-route models postulate that complex forms can be processed both as whole words and through morphological decomposition. One such model is the Augmented Addressed Morphology model (AAM, Caramazza, Laudanna, & Romani, 1988), according to which words are handled primarily by the direct route and the parsing route is a backup option for very rare or novel morphologically complex words. Another dual-route model is that of Schreuder and Baayen (1995), which has been applied to different languages such as Italian (Baayen, Burani & Schreuder, 1996) and Dutch (Baayen, Dijkstra & Schreuder, 1997). This model is named the Parallel Dual-Route model (PDR). In this model, storage and decomposition routes are activated in parallel, and their relative contributions to the recognition of morphologically complex words depend on a number of factors. Bertram, Schreuder, and Baayen (2000) proposed three main factors: Word formation type, suffix productivity, and affixal homonymy. Word formation type refers to the meaning relationship between the morphologically complex word and the base word (this is a continuum going from inflections that do not alter the meaning of the stem to derivations that have a different meaning than the stem word). The productivity of a suffix refers to the ease with which

This article was published Online First December 7, 2015.

Manuel Gimenes, Laboratoire Centre de Recherches sur la Cognition et l'Apprentissage, University of Poitiers, CNRS; Marc Brysbaert, Department of Experimental Psychology, Ghent University; Boris New, Laboratoire de Psychologie et Neurocognition, CNRS, University Savoie Mont Blanc.

We thank Marcus Taft, Matt Davis, and Harald Baayen for their helpful comments on previous versions of this article. We would also like to thank Emmanuel Keuleers for his help with Celex.

Correspondence concerning this article should be addressed to Manuel Gimenes, Laboratoire Centre de Recherches sur la Cognition et l'Apprentissage, CNRS, UMR 7295, MSHS-Bâtiment A5, 5 rue Théodore Lefebvre, 86000 Poitiers, France. E-mail: manuel.gimenes@univ-poitiers.fr

new words can be created by adding the suffix. Finally, the third factor (affixal homonymy) distinguishes affixes that are used in only one form of derivation/inflection, versus affixes that are used in two or more types of derivation/inflection. According to Bertram et al. (2000), the contribution of the decomposition route is strongest for words with productive meaning-invariant affixes that do not have productive semantic rival homonyms. In contrast, whole-word recognition is more likely for affixes that are not productive or that have a more frequent rival with a different semantic function.

Connectionist models have also been proposed to account for plural processing. These models do not make a difference between regular words (which can be processed by rules) and exceptions. The weight of the connections between the different units can be modified and these changes allow the network to acquire both rules and exceptions. For example, Davis, van Casteren, and Marslen-Wilson (2003) tried to account for the singular and plural word processing results from Baayen et al. (1997) with a 3-layer model: One layer of orthographic units, one layer of hidden units, and one layer of semantic units (attributed randomly). Interestingly, Davis et al. (2003) could simulate Baayen's results only when they introduced affix homonymy in the simulation (more precisely, the authors had to take into account that some noun forms in the stimulus list also existed as verb forms). More recent computational models have been proposed (i.e., Baayen, Milin, Filipović, Hendrix, & Marelli, 2011) but they have difficulties in resolving the meaning of ambiguous affixes that attach to both nouns and verbs and have different meanings in the two cases. This is probably because these models only have a single layer of weights, without intervening hidden unit representations, and hence cannot solve nonlinearly separable mappings like XOR, or (equivalently) determining whether a word like "axes" is a noun plural or a third person singular verb, or indeed both.

Singular and Plural Nouns in French, English, and Dutch

In the framework of the three factors mentioned by Bertram et al. (2000), the study of plural nouns is particularly interesting because adding the suffix *-s* to a noun is an extremely productive way of forming plurals and it does not dramatically change the meaning of the stem word. Another interesting feature of the plural form is that the same suffix *-s* is used in three languages with different inflectional morphology: French, English, and Dutch. In English this ending is quite frequent and also used as the third person singular present. In French, *-s* is often used as a part of verbal affixes (such as *-es* [tu joues], *-ais* [tu jouais], *-ons* [nous marchons]) but is very rarely used as a verbal affix per se. In Dutch, the pattern is somewhat more complex, because there are two plural forms: *-en* and *-s*. Although *-en* is the most common plural affix in Dutch, the alternative affix *-s* also applies to a significant proportion of the nouns (31 % according to Baayen, McQueen, Dijkstra, & Schreuder, 2003) and does not have a higher-frequency competitor. In addition, many Dutch speakers also read English and French, meaning that they are quite familiar with the *-s* plurals (for morphological parsing in second language processing, see Diependaele, Duñabeitia, Morris, & Keuleers, 2011).

Surface and Base Frequencies

The different ways in which plural nouns can be processed have implications for the calculation of word frequency estimates. If plural and singular noun forms have their own lexical representations, as defended by full listing models, then the surface frequency of each form is the best frequency estimate. The surface frequency of a word form is the token frequency (per million) with which this particular word form appears in a representative corpus. For instance, the SUBTLEX surface frequency of the word *dog* is 193 per million (pm) words; that of the word *dogs* is 53 pm (Brysbaert & New, 2009). On the other hand, according to full parsing models like Taft (1994), not only the surface frequencies are what matters but also the base frequency. This is the sum of the frequencies of all inflections of a word, also sometimes called the lemma frequency. The base frequency is 246 pm both for *dog* and *dogs*.

Following Taft's model, for a singular noun the base frequency would be more important than the surface frequency: A singular noun is activated not only when a singular word form is presented but also when a plural noun form is presented (i.e., the lexical representation of *dog* is activated both when the words *dog* and *dogs* are presented). The PDR model makes the same prediction and the mechanism by which singulars are influenced by base frequency is the same. Conversely, full listing models predict a surface frequency effect for singular nouns.

For plural nouns, Taft's model predicts both surface and base frequency effects. Because *dogs* is first decomposed as *dog* + *-s*, the base frequency influences the speed with which word stems will be processed. Moreover, the surface frequency of *dogs* can also influence the recognition duration, but at a later stage of processing, as the time to decide whether the stem and the suffix can be combined depends of the surface frequency of the combination. Full listing models predict surface frequency effects but no base frequency effects. The parallel dual-route (PDR) model predicts that surface and base frequencies influence reaction times (RTs) to a certain extent. More precisely, according to the PDR model, neither surface frequency nor base frequency will be the overall "best" frequency estimate; their contributions will depend on the relative frequencies of the word forms involved and on the type of word presented. For a plural noun, depending on its frequency, it could be decomposed or stored. For a low-frequency plural, it will be decomposed. This will imply the activation of its singular, which is sensitive to base frequency. As a result, for low-frequency plural noun forms the base frequency is expected to be more important than the surface frequency. Taft's model makes the same overall prediction (except for one difference, which we will describe below).

A prediction of the PDR model and of Taft's model, therefore, is that base frequency will be more informative to predict processing times of singular nouns, whereas surface frequency or base frequency will have different influences depending on plural nouns' frequencies. This prediction was confirmed in a number of studies. In a first experiment, Baayen et al. (1997) kept the base frequency of the stimuli constant and manipulated the surface frequency. The results showed that lexical decision times to the plural forms were influenced by surface frequency. In contrast, lexical decision times to the singular forms did not show any surface frequency effect. In a second experiment, Baayen et al.

kept the surface frequency of the singular nouns constant, but manipulated the frequencies of the plural forms. They observed a significant effect of the base frequency on the lexical decision times to the singular word forms. Importantly for our study, Baayen et al. (2003) showed that the same pattern of results was obtained in Dutch with plurals ending in *-en* and with plurals ending in *-s*. New, Brylsbaert, Segui, Ferrand, and Rastle (2004) replicated the pattern in French. They found very similar results in English (New et al., 2004; Sereno & Jongman, 1997) with only a slight difference. New et al. (2004) hypothesised that this slight deviation was due to the fact that many noun forms in English also exist as verb forms (*play*, *plays*).

In sum, the results of Baayen et al. (1997, 2003) and New et al. (2004) suggest that response times to singular nouns are best predicted by base frequency, whereas response times to plural nouns are predicted by surface frequency. Another interesting result was found by Taft (2004): The base frequency effect is sensitive to the type of nonwords used in the lexical-decision task. When the recombination stage was not crucial for distinguishing the words from the nonwords (e.g., when the nonwords were inflected nonsense stems, like “milphs” or “juxing”), a base frequency effect was observed (and also a weak but significant surface frequency effect). However, when the recombination stage was critical for differentiating the words from the nonwords (e.g., when the nonwords had a real word stem, like “mirths” or “joying”), a reverse base frequency effect was observed. Taft (2004) argued that the PDR model was not able to explain such a reverse base frequency effect (the PDR model would predict a surface frequency effect and a lack of base frequency effect, but not a reverse base frequency effect).

None of these studies compared in the same experiment the influence of both surface and base frequency effects (when one of these was manipulated the other was controlled and vice versa). One of the goals of the current paper is to use the same experiment to compare both surface and base frequency effects because the PDR and Taft models do not predict exactly the same effects (see the next section).

Goals of the Study

The first goal of this study was to explore whether the processing of singular and plural is the same in French, English, and Dutch, as suggested by the experiments described above. Given the language differences in morphology, this would indicate that the pattern is very robust.

A second goal of our study was to test precise predictions from two models of morphological processing (Taft’s model and the PDR model). First, for the three languages, as plural forms should activate singular whenever they are processed, base frequency should be the only predictor for singulars (no effect of surface frequency). This prediction is made both by Taft’s model and the PDR model. Second, low-frequency plurals should be decomposed. So, we can hypothesise that we will observe a strong base frequency effect for these forms according to Taft and PDR models. The Taft model also predicts an effect of surface frequency (concomitant with the effect of base frequency) because of the late stage of recombination, while the PDR does not. Third, as high-frequency plural forms are stored, they should show a surface-frequency effect. Finally, the PDR and Taft models do not

make the same prediction for the medium-frequency plurals. According to Taft’s model, medium-frequency plurals should show a base frequency effect (because of the first stage of decomposition) and also a surface frequency effect (because of the late stage of recombination). According to the PDR model, every word should be recognised either through decomposition or through whole-word access (this is determined by the nature of the suffix and the surface frequency of the word). Therefore, the PDR model predicts either a surface effect or a base effect for medium-frequency plurals, but not both effects for a single word.

A third goal of our study was to test the predictions of the PDR and Taft’s models with a correlational approach, rather than the traditional factorial approach. The use of regression analyses allows us to directly compare effects of surface and base frequency, both for singular and plural forms, and across a much wider range of values. Such comparisons are very difficult to realise in factorial experiments (see Brylsbaert, Stevens, Mandler, & Keuleers, *in press*, for a more in-depth discussion). Some authors have recommended the use of correlational designs in addition to factorial designs (Ford, Marslen-Wilson, & Davis, 2003; Balota, Yap, Hutchison, & Cortese, 2012; Sibley, Kello, & Seidenberg, 2009).

The Use of Megastudy Data

For correlational studies we need more data than for a typical factorial design. The availability of megastudy data is a good solution to this problem. Megastudies are large databases of descriptive and behavioural data. The first such study was published by Balota et al. (2007). In the so-called English Lexicon Project (ELP) they collected naming times and lexical decision times for over 40,000 English words from several hundred participants. The lexical decision approach was repeated in the French Lexicon Project (FLP; Ferrand et al., 2010), the British Lexicon Project (BLP; Keuleers, Lacey, Rastle, & Brylsbaert, 2012), the Dutch Lexicon Project (DLP; Keuleers, Diependaele, & Brylsbaert, 2010), and the Malay Lexicon Project (Yap, Liow, Jalil, & Faizal, 2010).

In this article, we used data from three megastudies: FLP, BLP, and DLP. We did not use ELP because the nonwords in that study were built with a different procedure. In ELP, nonwords were generated by changing one letter in a corresponding target word. This way of constructing nonwords confounded nonword length and word likeness (New, Ferrand, Pallier, & Brylsbaert, 2006). In contrast, in FLP, BLP, and DLP, nonwords were generated in such a way that their orthographic features mimicked the orthographic features of the words. The main characteristics of these three megastudies are presented in Table 1.

Correlational Analyses

We used the same selection criteria for the three megastudy databases. First, we selected the nouns with a regular *-s* plural. Second, we discarded the words that also existed as verbs or adjectives. This may have led to the selection of a slightly different subset of English words compared to the other two languages, given that in English a substantial set of words can be used both as noun and as verb. Nevertheless, we made this decision because the fact that some words can exist as nouns and as verbs makes the test of our hypotheses much more complicated. Third, for all languages

Table 1
Main Features of Three Megastudies

	FLP	BLP	DLP
Number of words	38,840	28,730	14,089
Number of nonwords	38,840	28,730	14,089
Method to create nonwords	See article	Wuggy	Wuggy (Keuleers & Brysbaert, 2010)
Number of words for each participant	1,000	Half (14,365)	All
Number of syllables	1 to 6 syllables	Monosyl. Disyl.	Monosyl. Disyl.
Number of participants	975	78	39

Note. FLP = French Lexicon Project; BLP = British Lexicon Project; DLP = Dutch Lexicon Project; Monosyl. = monosyllabic; Disyl. = disyllabic.

we used frequency estimates based on a subtitle corpus. In previous analyses these have been shown to be better predictors of lexical decision times than frequencies based on written sources (New, Brysbaert, Veronis & Pallier, 2007 for French; Brysbaert & New, 2009, and Brysbaert, Keuleers, & New, 2011 for English; Keuleers, Brysbaert, & New, 2010 for Dutch). In addition, because the subtitle corpora were compiled on very much the same sources (popular films and TV series), the estimates are comparable in the different languages.

For each singular and plural noun, we had the lexical decision time, the surface frequency and the base frequency (defined as the sum of the singular and the plural forms). These were entered in regression models with item-averaged RT as the dependent variable and surface and/or base frequency as predictors. To control for confounding variables, the regression analyses also included the following independent variables (computed from the singular form): number of letters in the word, number of syllables, and the orthographic neighborhood measure OLD20 (Yarkoni, Balota, & Yap, 2008). The frequency measures were log transformed, and we used a polynomial of the third degree, as it has been shown that the frequency effect is nonlinear (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Baayen, Feldman, & Schreuder, 2006; Keuleers, Brysbaert et al., 2010). We only kept words with lengths between 3 and 13 letters, frequencies higher than 0, and accuracies higher than .67 (to make sure the RTs were based on words known by the majority of the participants).

Because of our predictions, we looked at selected parts of the frequency range to get a more detailed idea of what is happening at the low and the high end of the continuum. Based on the frequency effects observed in French, English, and Dutch in Keuleers, Brysbaert et al. (2010), we chose a threshold for our high-frequency words of 20 pm words ($\log_{10} = 1.3$) as we can see in their Figure 4 that the frequency effect is having an inflection point around that value. For the low-frequency threshold we chose 2 in order to have a large gap (Factor 10) between our high- and low-frequency items. That allowed us to run analyses for the words with surface frequencies below 2 pm (low-frequency words), frequencies between 2 pm and 20 pm (medium-frequency words), and frequencies above 20 pm (high-frequency words).

In this study, we need to compare the effects of surface and base frequencies, which are highly correlated. This raises the issue of collinearity (i.e., the fact that it is problematic to include correlated predictor variables in a simultaneous regression model; Baayen, 2010). However, collinearity particularly affects the regression weights, not the total percentage of variance explained (Freund, Wilson, & Sa, 2006). So, a straightforward way out of the problem

is not to look at the significance of the regression weights, but at the percentage of variance accounted for. We did so by building models with only one frequency measure, then with both, which were then compared using *F* tests. In other words, we always ran three analyses: one with surface frequency, one with base frequency, and one with both frequencies. We used by-item Anova tests to determine if the part of variance uniquely explained by base frequency or surface frequency was significant. When they were both significant, we also looked at which one had the best fit. A second way to address the collinearity issue is to decrease the correlation between the predictors. This can be done by not using $\log(\text{surface frequency})$ but $\log(\text{surface frequency}/\text{base frequency})$; del Prado Martín, Kostić, & Baayen, 2004). This measure has been used to study the Dutch plural data and found to be a significant predictor of visual lexical decision times (Baayen, Levelt, Schreuder, & Ernestus, 2008). For the next parts of this article, we call this measure “Surface Information.” This measure reduces the correlation with base frequency and addresses a second potential problem from a regression perspective, namely the fact that the surface frequency could be considered as part of the base frequency measure.

Regression of the FLP Data

In French, the criteria described above resulted in the selection of 2,954 words. We used Lexique 3.70 (subtitle) as corpus frequency (New et al., 2007). The results of the regression analyses are presented in Table 2.

Concerning our new measure, Surface Information, we observed that on average its correlation with base frequency (0.4) was less than the original surface frequency measure (0.57). For this reason, when we examine the unique impact of base and surface frequency, we will focus on the analysis using the Surface Information measure. However, it can be noted that very similar results were obtained with surface frequencies. We will also follow this approach to describe British and Dutch results.

For singular forms, whatever their frequency range, base frequency is what matters. This agrees with the factorial designs and indicates that singular forms are coactivated when the plural form is processed, as predicted by the PDR model and the Taft model.

For plural forms, the effects of word frequency were not significant for frequencies above 20 pm. This is partly due to the small numbers of stimuli in this category, but is also in line with the recent observation that the frequency effect reaches a floor at the high end of the frequency continuum (Keuleers, Brysbaert et al., 2010). In megastudies, the frequency effect is next to nonexistent

Table 2

Regression Results in French Lexicon Project for Singular and Plural Nouns and for Each Frequency Range

	Singular nouns			Plural nouns		
	Freq > 20	Freq < 20 and > 2	Freq < 2	Freq > 20	Freq < 20 and > 2	Freq < 2
Nb words	521	1461	972	129	942	1884
Surface and Base						
Correlation Surface–Base	.97	.68	.46	.79	.30	.23
R^2 Surface	.160	.148	.158	.101	.227	.209
R^2 Base	.158	.162	.192	.092	.224	.31
R^2 Surface and Base	.164	.164	.194	.106	.245	.314
Surface unique	ns	ns	ns	ns	***	*
Base unique	ns	***	***	ns	***	***
Surface Info and Base						
Correlation Surface Info–Base	.04	.45	.38	.45	.62	.47
Surface Info unique	ns	ns	ns	ns	***	*
Base unique	**	***	***	ns	***	***

Note. ns = non significant.

* $p < .05$. ** $p < .01$. *** $p < .001$.

for frequencies above 50 pm Medium frequency plurals showed both surface and base frequency effects (as predicted by Taft's model): The two frequencies explained the same amount of variance. Finally, for low frequency plurals, both surface frequency and base frequency matter, but base frequency explains more variance than surface frequency. This last result was not predicted by the PDR and Taft models.

Regression of the BLP Data

We used the same criteria as for the FLP study, which resulted in the selection of 1,475 words. Frequencies were based on SUBTLEX-US (Bryspaert & New, 2009). Table 3 shows the results of the regression analyses.

For singular forms, the results were similar to French: Whatever their frequency range, base frequency is what matters. There is only one difference with French: For medium frequencies in British, both surface and base frequencies were significant.

For plural forms, there were no significant effects above 20 pm. Finally, for medium-frequency plurals and for low-frequency plu-

als, while base frequency explained more variance than surface frequency, both were significant.

In summary, the English data fully agree with those of FLP, as could be expected from the similar formation of plural nouns in English and French. More interesting is to see whether the pattern generalizes to Dutch.

Regression of the DLP Data

The same word selection criteria were used, resulting in 544 words. Word frequencies were based on the SUBTLEX-NL corpus (Keuleers, Bryspaert et al., 2010). Table 4 shows the results of the regression analyses. When interpreting these data, it is important to keep in mind that the number of stimuli was considerably smaller, putting more severe limits on what can be observed.

For singular forms, the results were the same as those in French and English (base frequency effect but no surface information frequency effect), except for one aspect: There was no base frequency effect for high-frequency singulars. However, the small number of words could lead to a lack of power.

Table 3

Regression Results in British Lexicon Project for Singular and Plural Nouns and for Each Frequency Range

	Singular nouns			Plural nouns		
	Freq > 20	Freq < 20 and > 2	Freq < 2	Freq > 20	Freq < 20 and > 2	Freq < 2
Nb words	251	657	567	69	424	984
Surface and Base						
Correlation Surface–Base	.96	.62	.48	.76	.29	.14
R^2 Surface	.087	.209	.158	.157	.155	.203
R^2 Base	.087	.21	.202	.136	.171	.227
R^2 Surface and Base	.091	.221	.205	.168	.204	.261
Surface unique	ns	*	ns	ns	***	***
Base unique	ns	*	***	ns	***	***
Surface Info and Base						
Correlation Surface Info–Base	.16	.51	.35	.37	.65	.34
Surface Info unique	ns	*	ns	ns	***	***
Base unique	*	***	***	ns	***	***

Note. ns = non significant.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4

Regression Results in Dutch Lexicon Project for Singular and Plural Nouns and for Each Frequency Range

	Singular nouns			Plural nouns		
	Freq > 20	Freq < 20 and > 2	Freq < 2	Freq > 20	Freq < 20 and > 2	Freq < 2
Nb words	81	252	211	17	127	399
Surface and Base						
Correlation Surface–Base	.90	.68	.65	.86	.38	.07
R^2 Surface	.167	.138	.073	.597	.107	.111
R^2 Base	.171	.153	.14	.450	.074	.118
R^2 Surface and Base	.195	.157	.153	.610	.147	.151
Surface unique	ns	ns	ns	ns	*	**
Base unique	ns	ns	***	ns	ns	***
Surface Info and Base						
Correlation Surface Info–Base	.17	.63	.27	.28	.69	.48
Surface Info unique	ns	ns	ns	ns	*	**
Base unique	ns	***	***	ns	**	***

Note. ns = non significant.

* $p < .05$. ** $p < .01$. *** $p < .001$.

For plural forms, the pattern was the same as in French and English: both significant surface and base frequency effects for medium-frequency and low-frequency plurals but not for high-frequency plurals.

In the three languages, we observed effects of both base frequency and surface frequency for low-frequency plural words. This result runs against the PDR model as it predicts that the base frequency should be the only one that matters for less frequent words. A possible reason for this contradiction could be that a threshold of 2 pm words is too high. Therefore, for each language we searched for the highest threshold X that would agree with the predictions of the PDR model. In order to do this, we compared the unique influence of base and surface frequencies on a subset of plurals having a surface frequency lower than X. We repeated the regressions by decreasing X to see whether for a certain X value (our threshold) it was possible to find only a base frequency effect. In order to get a distribution of the thresholds for the different languages we ran a bootstrap analysis. For each bootstrap we took a random sample of the same size as the full dataset but sampling words at random with replacement and we did this 100 times. The variation in the frequency threshold between these different bootstraps gave us a distribution of thresholds for the different languages which allows some statistical comparison. This allowed us to create 95% confidence intervals for the estimated threshold in each language. The results are shown in Table 5.

As can be seen in Table 5, and as predicted by the PDR model, it was possible to find a threshold below which only a base frequency effect (and no surface frequency effect) was observed. However, the location of the threshold was significantly different in the three languages as the 95 % confidence intervals did not overlap: We observed the highest threshold for French (1.77), a lower one in Dutch (0.84), and the lowest in English (0.71). This cross-linguistic difference will be discussed below.

Discussion

In this study, we explored the processing of singular and plural noun forms using recently collected megastudy data for French, English, and Dutch. This allowed us to analyse larger sets of stimuli with regression analyses (Tables 2–5). Concerning singu-

lars, our findings confirmed that base frequency (i.e., the combined frequency of the singular and the plural form) is more informative to predict lexical decision times than surface frequency. For the low-frequency plural forms in the three languages, we observed effects of both base frequency and surface frequency. Only for very low-frequency words did we observe a single effect of base frequency, but with different thresholds in the three languages. In the following, we will discuss these findings in the framework of the three models presented in the introduction.

Concerning the singulars, a strong prediction both from Taft (1994, 2004) and from the PDR model is that whenever a plural form is presented, it should coactivate the singular form as the singular form is physically part of the plural form. Consistent with this prediction, we observed that singular forms were influenced by their base frequency, but not by their unique surface frequency. This result is particularly robust as we observed it in the three languages and in general for low-, medium-, and high-frequency singulars. Interestingly, Longtin, Segui, and Hallé (2003) and Rastle, Davis, and New (2004) drew similar conclusions using a completely different methodology. Using lexical decision with

Table 5

Thresholds (With 95% Confidence Intervals) for Very-Low-Frequency Plurals and Regression Results in FLP, BLP, and DLP

	Plural nouns		
	FLP	BLP	DLP
Whole data			
Plural frequency Threshold	1.77	.71	.84
Nb words	1774	637	312
Correlation Surface Info–Base	.48***	.30***	.49***
Surface Info unique	ns	ns	ns
Base unique	***	***	***
Bootstrapping			
95% CI Lower bound	1.51	.61	.80
95% CI Upper bound	1.83	.73	1.03

Note. FLP = French Lexicon Project; BLP = British Lexicon Project; DLP = Dutch Lexicon Project; ns = non significant.

* $p < .05$. ** $p < .01$. *** $p < .001$.

masked priming, they showed that morphologically complex primes facilitate target processing (*corner-CORN*) while orthographic controls do not (*brothel-BROTH*). They concluded that an automatic process of decomposition is applied to all morphologically complex stimuli (but see Baayen et al., 2011 for another interpretation of this result). More recently, it has been shown that the automatic morphological decomposition also applies for high-frequency words (McCormick, Brylsbaert & Rastle, 2009). This is completely coherent with the finding in the present paper that whenever a plural is presented, it is automatically decomposed in its singular and its suffix. Although this result was predicted by Taft (1994, 2004) and the PDR model, this is not compatible with AAM and a full-listing model.

The fact that the base frequency is so important for singulars also has methodological implications because it strongly suggests that the best frequency estimate for singular nouns is the base frequency (or lemma frequency) rather than the surface frequency. This is important, as the advice given thus far was to work with surface frequencies for all word forms (e.g., Brylsbaert & New, 2009; Brylsbaert et al., 2011). Indeed, word frequency is the most important variable in word processing and a suboptimal measure can easily bias other variables.

Concerning the high-frequency plurals, there were neither significant effects of unique base nor unique surface frequency but these are also the analyses where we had the least items, which could explain the lack of significance. This is also in line with the strong nonlinearity of the frequency effect (Baayen, 2005; Baayen et al., 2006; Balota et al., 2004; Keuleers, Diependaele et al., 2010) and the observation that the frequency effect is much weaker for high than for low-frequency word forms. However, the tendency in the three languages was in favour of the surface frequency for the plural form (as can be seen with the R^2 values), as predicted by the PDR and Taft models (but this is speculation).

Concerning the medium and the low-frequency plurals we observed both a base frequency and a surface frequency effect in all three languages. This cannot be explained by the full listing or AAM models as they predict mainly a surface frequency effect. However, this finding is consistent with Taft's model, where each word is processed at two levels: an early stage of decomposition (influenced by the base frequency) and a late stage of recomposition (influenced by the surface frequency). The presence of both surface and base effects is difficult to explain for the PDR model. Indeed, every word should be recognised mainly through decomposition or through whole-word access (the chosen route is determined by the type of suffix and the surface frequency of the word). Therefore, it is difficult for the PDR model to explain how both surface and base frequency can have an influence on the same set of words. A possible explanation could be that in this range there are two different types of words: One type mainly processed by decomposition, and another stored in the lexicon. Further research is needed to see whether it is possible to find a parameter set that would distinguish between these two types of words.

A search with lower thresholds showed that base frequency was the most important variable predicting lexical decision times for words with frequencies lower than 1.77 pm in French, .84 in Dutch, and .71 in English, but below 2 pm for all the three languages. This "base frequency effect only" effect for very low-frequency plural nouns was predicted by the PDR model. It was not predicted by Taft's model, but could be explained within the

model by assuming that if participants have spent a long time activating the lemma of the stem (because of its low frequency) they respond without bothering to establish (at the recombination stage) if the plural is possible. For the three languages, the threshold was substantially lower than the threshold of 6 pm proposed by Alegre and Gordon (1999). The latter study was criticised by Baayen, Wurm, and Aycock (2007), who found that a surface frequency effect can be obtained with items below 6 pm if given sufficient power. The study was also based on the Francis and Kucera (1982) frequency counts, which does not allow precise frequency estimates below 6 pm (Brylsbaert & New, 2009). Indeed, only since the introduction of large corpora has it become clear that over half of the word forms have frequencies below 1 pm (van Heuven, Mandera, Keuleers, & Brylsbaert, 2014).

The cause of the difference between the three languages remains to be explored but we can formulate one hypothesis based on the inflection system used by the three languages. This hierarchy of thresholds between French, Dutch, and English could be explained by the fact that there are many more inflected words in French than in Dutch, and in Dutch than in English. For instance, in French, verbs are conjugated to reflect the person, the mood, the tense, and thus can have as many as 30 different forms. In Dutch, there are fewer suffixes, but there are about 10 possible suffixes to conjugate verbs, for example. In English, there are much less inflected forms (no gender and only 3 inflected forms for verbs). So, if we had found similar thresholds in the three languages, that would have meant that French was storing many more complex forms than Dutch and English. For example, if we count the number of inflected words in the SUBTLEX corpus with a frequency less than 1.77 pm we find in French 87,411 words, whereas in English there are only 26,917 such forms (3 times less). If this explanation is true, it means that the inflection richness of a language may be another factor explaining the relative contribution of storage versus decomposition.

The next step now is to see what our findings imply for other languages or other word types. For example, Spanish is less morphologically rich than French but more than English. Therefore, we can predict that the Spanish threshold (below which only base frequency predicts lexical decision times) should be between the French threshold (1.77) and the English (0.71). Our results also raise questions about verb morphology: Do our results mean that processing times of verb stems (e.g., *eat*) are also better predicted by base frequency, whereas surface frequency is what matters most for inflected forms above the critical threshold? These are questions that are still ahead of us.

To sum up, we used megastudies to better understand the cognitive processing of singular and plural noun processing. We found that singular nouns are sensitive to base frequency. This result shows that the singular form is always activated during plural form processing, which is compatible with the hypothesis of automatic, prelexical morphological decomposition (Longtin et al., 2003; Rastle et al., 2004). We also found that medium- and low-frequency plural nouns are sensitive to both surface and base frequencies, which suggests a mixture of both decomposed and whole word contributions. Finally, for very low-frequency words, we found only a base frequency effect. This indicates that below a given threshold, morphological decomposition is systematically used to process plural forms. Interestingly, the threshold separating low- from very-low-frequency words was different in the three

languages. This cross-linguistic difference could be explained by the morphological richness of the language. A morphological rich language such as French would involve more decomposition than English, but this hypothesis needs future research to be confirmed.

Résumé

Dans cette étude, nous explorons le traitement des formes lexicales au singulier et au pluriel, au moyen de méga-études en français, en anglais et en néerlandais. En ce qui a trait au singulier, nous avons constaté un effet de la fréquence de base, mais aucun effet de la fréquence de surface. Quant aux pluriels, l'effet dépendait de la fréquence de la forme lexicale. Quand la forme lexicale avait une fréquence supérieure à une valeur-seuil, on a constaté un effet de la fréquence de base et de surface. En ce qui a trait aux fréquences inférieures au seuil, on a constaté un effet de la fréquence de base seulement, ce qui laisse supposer une décomposition entière de ces mots. Le seuil variait selon la langue, ce qui laisse croire qu'un plus grand nombre de mots pluriels sont décomposés en français, comparativement au néerlandais, et un plus grand nombre en néerlandais qu'en anglais. Toutefois, pour les trois langues, il semble y avoir coactivation du singulier chaque fois que la forme au pluriel est traitée. Ces résultats sont ensuite interprétés en fonction de certains des principaux modèles du traitement morphologique.

Mots-clés : reconnaissance visuelle des mots, morphologie, pluriel, méga-études.

References

- Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40, 41–61. <http://dx.doi.org/10.1006/jmla.1998.2607>
- Baayen, R. H. (2005). Data mining at the intersection of psychology and linguistics. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 69–83). Mahwah, NJ: Erlbaum.
- Baayen, R. H. (2010). A real experiment is a factorial experiment? *The Mental Lexicon*, 5, 149–157. <http://dx.doi.org/10.1075/ml.5.1.06baa>
- Baayen, R. H., Burani, C., & Schreuder, R. (1996). Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian. In G. E. Booij & J. van Marle (Eds.), *Yearbook of morphology 1996* (pp. 13–33). Dordrecht, The Netherlands: Kluwer Academic.
- Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, 37, 94–117. <http://dx.doi.org/10.1006/jmla.1997.2509>
- Baayen, R. H., Feldman, L. F., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55, 290–313. <http://dx.doi.org/10.1016/j.jml.2006.03.008>
- Baayen, R. H., Levelt, W. M. J., Schreuder, R., & Ernestus, M. (2008). Paradigmatic structure in speech production. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, 43, 1–29.
- Baayen, R. H., McQueen, J. M., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 355–390). Berlin, Germany: De Gruyter Mouton. <http://dx.doi.org/10.1515/9783110910186.355>
- Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118, 438–481. <http://dx.doi.org/10.1037/a0023851>
- Baayen, R. H., Wurm, L. H., & Ayccock, J. (2007). Lexical dynamics for low frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, 2, 419–463. <http://dx.doi.org/10.1075/ml.2.3.06baa>
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133, 283–316. <http://dx.doi.org/10.1037/0096-3445.133.2.283>
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459. <http://dx.doi.org/10.3758/BF03193014>
- Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: Large scale analysis of lexical processes. In J. S. Adelman (Ed.), *Visual word recognition* (Vol. 1, pp. 90–115). London, UK: Psychology Press.
- Bertram, R., Schreuder, R., & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 489–511. <http://dx.doi.org/10.1037/0278-7393.26.2.489>
- Brysbaert, M., Keuleers, E., & New, B. (2011). Assessing the usefulness of google books' word frequencies for psycholinguistic research on word processing. *Frontiers in Psychology*, 2, 27. <http://dx.doi.org/10.3389/fpsyg.2011.00027>
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977–990. <http://dx.doi.org/10.3758/BRM.41.4.977>
- Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (in press). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*.
- Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production Vol. 2: Development, writing and other language processes* (pp. 257–294). London, UK: Academic Press.
- Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28, 297–332. [http://dx.doi.org/10.1016/0010-0277\(88\)90017-0](http://dx.doi.org/10.1016/0010-0277(88)90017-0)
- Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22, 991–1013. <http://dx.doi.org/10.1017/S0140525X99002228>
- Davis, M. H., van Casteren, M., & Marslen-Wilson, W. D. (2003). Frequency effects in processing inflected Dutch nouns: A distributed connectionist account. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 427–462). Berlin, Germany: De Gruyter Mouton. <http://dx.doi.org/10.1515/9783110910186.427>
- del Prado Martín, F. M., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94, 1–18. <http://dx.doi.org/10.1016/j.cognition.2003.10.015>
- Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, 64, 344–358. <http://dx.doi.org/10.1016/j.jml.2011.01.003>
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., . . . Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42, 488–496. <http://dx.doi.org/10.3758/BRM.42.2.488>
- Ford, M. A., Marslen-Wilson, W. D., & Davis, M. H. (2003). Morphology and frequency: Contrasting methodologies. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp.

- 89–124). Berlin: De Gruyter Mouton. <http://dx.doi.org/10.1515/9783110910186.89>
- Francis, W., & Kučera, H. (1982). *Frequency analysis of English language usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.
- Freund, R. J., Wilson, W. J., & Sa, P. (2006). *Regression analysis: Statistical modeling of a response variable*. London, UK: Academic Press.
- Keuleers, E., & Brylsbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42, 627–633.
- Keuleers, E., Brylsbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42, 643–650. <http://dx.doi.org/10.3758/BRM.42.3.643>
- Keuleers, E., Diependaele, K., & Brylsbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, 1, 174. <http://dx.doi.org/10.3389/fpsyg.2010.00174>
- Keuleers, E., Lacey, P., Rastle, K., & Brylsbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, 44, 287–304. <http://dx.doi.org/10.3758/s13428-011-0118-4>
- Longtin, C.-M., Segui, J., & Hallé, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, 18, 313–334. <http://dx.doi.org/10.1080/01690960244000036>
- Manelis, L., & Tharp, D. A. (1977). The processing of affixed words. *Memory & Cognition*, 5, 690–695. <http://dx.doi.org/10.3758/BF03197417>
- McCormick, S. F., Brylsbaert, M., & Rastle, K. (2009). Is morphological decomposition limited to low-frequency words? *Quarterly Journal of Experimental Psychology*, 62, 1706–1715.
- McCormick, S. F., Rastle, K., & Davis, M. H. (2008). Is there a ‘fete’ in ‘fetish’? Effects of orthographic opacity on morpho-orthographic segmentation in visual word recognition. *Journal of Memory and Language*, 58, 307–326. <http://dx.doi.org/10.1016/j.jml.2007.05.006>
- New, B., Brylsbaert, M., Segui, J., Ferrand, L., & Rastle, K. (2004). The processing of singular and plural nouns in French and English. *Journal of Memory and Language*, 51, 568–585. <http://dx.doi.org/10.1016/j.jml.2004.06.010>
- New, B., Brylsbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28, 661–677. <http://dx.doi.org/10.1017/S014271640707035X>
- New, B., Ferrand, L., Pallier, C., & Brylsbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, 13, 45–52. <http://dx.doi.org/10.3758/BF03193811>
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother’s brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11, 1090–1098. <http://dx.doi.org/10.3758/BF03196742>
- Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ: Erlbaum.
- Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition*, 25, 425–437. <http://dx.doi.org/10.3758/BF03201119>
- Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (2009). Error, error everywhere: A look at megastudies of word reading. In N. Taatgen & H. Van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1036–1041). Amsterdam, The Netherlands: Cognitive Science Society.
- Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, 9, 271–294. <http://dx.doi.org/10.1080/01690969408402120>
- Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology, A: Human Experimental Psychology*, 57, 745–765. <http://dx.doi.org/10.1080/02724980343000477>
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brylsbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67, 1176–1190. <http://dx.doi.org/10.1080/17470218.2013.850521>
- Yap, M. J., Liow, S. J. R., Jalil, S. B., & Faizal, S. S. B. (2010). The Malay Lexicon Project: A database of lexical statistics for 9,592 words. *Behavior Research Methods*, 42, 992–1003. <http://dx.doi.org/10.3758/BRM.42.4.992>
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart’s *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15, 971–979. <http://dx.doi.org/10.3758/PBR.15.5.971>

Received February 27, 2015

Accepted September 30, 2015 ■

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.