

Article

# Refunctionalization and Usage Frequency: An Exploratory Questionnaire Study

Malte Rosemeyer 

Romanisches Seminar, Albert-Ludwigs-Universität Freiburg, 79085 Freiburg, Germany; malte.rosemeyer@gmail.com

Received: 9 October 2018; Accepted: 19 October 2018; Published: 23 October 2018



**Abstract:** This paper explores the relationship between refunctionalization and usage frequency. In particular, it argues that (a) refunctionalization is more likely for low-frequency construction than high-frequency constructions, and that (b) high-frequency patterns are more likely candidates as models for refunctionalization processes than low-frequency patterns. It proposes that folk etymology processes be characterized as a type of refunctionalization process because in folk etymology, obsolescent and semantically void morphemes are replaced with morphemes that actually serve a function in language. This assumption allows for an empirical investigation of refunctionalization using an exploratory questionnaire study. The results indicate that usage frequency indeed plays a role in folk etymology processes, and consequently, refunctionalization. In particular, participants were more likely to accept false etymologies when the proposed etymon had a high usage frequency than when it had a low usage frequency. In summary, the present study proposes a way to study refunctionalization processes in synchrony.

**Keywords:** language change; historical linguistics; refunctionalization; frequency effects; folk etymology; Spanish

## 1. Introduction

Given that “*refunctionalization*”, the process by which a linguistic construction obtains a discourse-pragmatic function other than its original function, is a concept from historical linguistics, studies of refunctionalization normally adopt a diachronic perspective. Such diachronic studies usually identify the original discourse-pragmatic function of the construction, compare this original function to the function of the construction in later stages of the language in question, and hypothesize as to the reasons for this specific refunctionalization process.

In this paper, I explore the possibility of studying refunctionalization and the speakers’ motivation for such refunctionalization processes in synchrony. In particular, I investigate two hypotheses regarding refunctionalization derived from the standpoint of usage-based linguistics. First, it can be assumed that refunctionalization processes are more likely for low-frequency constructions than high-frequency constructions because in order for refunctionalization processes to become necessary, the construction in question has to become (near-)obsolescent. Second, when searching for models for the new function of the construction, speakers might rely to a greater degree on high-frequency patterns than on low-frequency patterns. I propose that folk etymology processes can be characterized as a type of refunctionalization process because in folk etymology, obsolescent and semantically void morphemes are replaced with morphemes that actually serve a function in language. This assumption allows for an empirical investigation of the motivation of refunctionalization processes using an exploratory questionnaire study. The questionnaire asked participants whether they would accept the assumption that a target word was derived from a proposed, false etymon. It therefore allowed analyzing the participants’ decision to accept or reject the folk etymology dependent of the usage

frequency of the target word and the proposed etymon, as well as other factors. The results from the questionnaire survey indicate that as in other refunctionalization processes, there is indeed a relationship between usage frequency and folk etymology, although not all of the assumptions are validated.

This paper is structured as follows. In Section 2, I discuss the relationship of usage frequency and refunctionalization processes, establishing the two hypotheses mentioned above. In Section 3, in turn, the relationship between refunctionalization and folk etymology is discussed. I propose that folk etymology be viewed as an instance of a process of refunctionalization. Section 4 describes the questionnaire survey conducted in order to study usage-based determinants of folk etymology. The results from the questionnaire survey are discussed in Section 5. The paper closes with a critical assessment of the findings and the methodology in the concluding Section 6.

## 2. Usage Frequency as a Determinant of Refunctionalization

In this paper, I will take the term “refunctionalization” to refer to the historical process by which a linguistic construction—a word, collocation, or grammatical pattern—obtains a discourse-pragmatic function other than its original function. As an example, consider the historical trajectory of the Swedish plural suffix *-on* (Norde 2009, pp. 181–83). As summarized in Norde and Van de Velde (2016, pp. 9–10), *-on* originally served as the nominative and accusative plural of neuter *an*-stems. Probably due to the fact that fruits such as *hiūpon* “rosehips” and *smultron* “wild” were usually referred to in the plural, the *-on*-suffix obtained a new function over time. In Modern Swedish, *-on* serves as a derivational suffix that is used to form fruit names such as *hallon* “raspberry” or *lingon* “lingonberry”. Note that such processes have been referred to using a multitude of terms (cf. Norde and Van de Velde 2016, pp. 10–11 for a summary). For instance, I could have easily used the related notion of “exaptation” (Lass 1990) in this paper in order to refer to this kind of historical mechanism.

The example of the Swedish *-on* suffix serves to illustrate the relevance of usage frequency for historical refunctionalization processes. First, there is an intrinsic connection between the usage frequency of the refunctionalized construction and the refunctionalization process. Norde and Van de Velde (2016, p. 10) note that Old Swedish *-on* is a relatively infrequent plural suffix, since most plurals were formed using a zero morpheme. In other words, refunctionalization appears to occur with obsolescent constructions, i.e., constructions that are being replaced by another construction in language change, a fact that has been frequently argued for in the literature (Lass 1990, pp. 81–82; Narrog 2016, pp. 98–99; Willis 2010, p. 151). In Rosemeyer (2014, p. 95), I argued that this effect is due to the nature of conserving effects in language change. It is well known that in processes of obsolescence in language change, instantiations of the replaced construction that are of relatively high usage frequency typically resist the change longer (Bybee 2006, 2010 et passim). As a result, obsolescing constructions are usually restricted to a very limited inventory of types. The low productivity of obsolescing constructions is necessarily correlated to the loss of their original grammatical function. For instance, consider the loss of subjunctive mood in Canadian French (Poplack 2001; Poplack et al. 2013). In comparison to European French, the subjunctive is used considerably less frequently in Canadian French. Especially in oral speech, the subjunctive is often replaced with the indicative (Poplack 2001, pp. 406–7). Poplack’s analysis of the determinants of the use of subjunctive mood in Canadian French reveals the existence of frequency asymmetries due to conserving effects. Thus, although the token frequency of the standard subjunctive variant is elevated, virtually all its uses are concentrated among a handful of highly favored matrix verbs collocated with a small cohort of frequent and irregular embedded verbs. Outside of these few contexts, in which its use has become ritualized, selection of the subjunctive is very rare. (Poplack 2001, p. 414)

This state of affairs involves the subjunctive no longer being used in all of the contexts in which its use would have been necessary given its original discourse-pragmatic function (the irrealis, i.e., the expression of “imagined, projected, predicted or otherwise unreal situations or events”, cf. Poplack 2001, p. 406). Given that there are cases in which the irrealis is expressed using the indicative

mood, in a sense the Canadian French subjunctive no longer fulfills the function of expressing the irrealis. Indeed, in a newer study, Poplack et al. (2013, p. 188) find that “any apparent semantic effect [i.e., irrealis, MR] was an epiphenomenon of the overriding effect of the lexical identity of the matrix”. It is this fracturing of the original discourse-pragmatic function that allows for refunctionalization; given the indecisiveness of speakers as to the semantics of obsolescing constructions, they might opt to assign these constructions a new function in discourse.

The relationship between usage frequency and refunctionalization processes described in the last paragraphs can be reformulated as Hypothesis 1 below.

**Hypothesis 1.** *Refunctionalization is more probable for low-frequency constructions than for high-frequency constructions.*

The usage frequency of the refunctionalized construction is not the only way in which frequency might affect the refunctionalization process. Rather, it seems plausible to assume that when speakers refunctionalize an obsolescing construction, they take grammatical patterns or functions of high usage frequency as models. As an example, consider auxiliary selection (*he comido* “I have eaten” vs. *soy venido* “I am come”) in Old and Early Modern Spanish, as analyzed in Rosemeyer (2014). Following a number of previous studies such as Mackenzie (2006) and Rodríguez Molina (2006), I conducted a quantitative analysis of the alternation in Old Spanish. This analysis suggests that even though *haber* “to have” and *ser* “to be” occur in comparable syntactic contexts (i.e., in frequent collocation with the past participle), they often have a different function. Whereas *haber* + PtcP has the temporal function of an anterior, *ser* + PtcP usually has a resultative function; it expresses the resultant state of a finished event. However, after the beginning of the 15th century *ser* + PtcP was gradually replaced by *haber* + PtcP, leading to its eventual demise in the 18th century. This obsolescence process leads to interesting changes in the use of *ser* + PtcP that suggest the existence of a refunctionalization process. In particular, the functional opposition of *haber* + PtcP and *ser* + PtcP in terms of the distinction between anterior and resultative constructions was gradually dissolved, with *ser* + PtcP being increasingly used as an anterior construction. Crucially, this refunctionalization process is not random; rather, *ser* + PtcP copied the discourse-pragmatic function of anteriority from the *haber* + PtcP construction that was very frequent in comparable usage contexts. In other words, the refunctionalization of *ser* + PtcP was a result from the actualization or analogical spread of *haber* + PtcP (see De Smet 2012 for a comprehensive discussion of actualization in language change).

As a second example, consider the changes in the paradigm of Romance oblique first- and second-person singular pronouns described in Smith (2006). In Latin, the opposition between the pronouns ME and MIHI > MI was one of case; whereas ME expressed accusative case, MIHI expressed dative case. However, due to the overall loss of the Latin case system in Proto-Romance, the two pronoun forms lost this function, and in some languages (Spanish, Northern French, and Italian dialects), came to be refunctionalized. In Modern Spanish, for instance, it is possible to use pronouns of both types in one sentence with the same reference and the same case. In (1a) *me* and (a) *mí* refer to the same referent and are datives. The difference between the two pronouns resides in the fact that whereas *me* is unstressed, *a mí* is stressed. The function of *a mí* is “disjunctive” in the sense that it serves as a focus expression; it eliminates all other possible candidates that might have been addressed. In this sense, strong personal pronouns such as *a mí* are no longer marked for case, as evident from the fact that whereas the use of a weak pronoun alone is acceptable (1b), the use of a strong pronoun alone is considered to be ungrammatical in most cases (Real Academia Española 2010, pp. 319–20) (1c).

(1)

- a.     *Me*        *lo*                *ha*                *dicho*        *a*        *mí.*  
to.me     it                have.PRS.3SG   say.PTCP   to        me  
“She said it to me (not you).”
- b.     *Me*        *lo*                *ha*                *dicho.*  
to.me     it                have.PRS.3SG   say.PTCP  
“She said it to me.”
- c.     ?*Lo*        *ha*                *dicho*        *a*        *mí.*  
it         have.PRS.3SG   say.PTCP   to        me  
“She said it to me (not you).”

It is possible that not only the usage frequency of the refunctionalized forms, but also the usage frequency of the function they came to adapt had an influence on the realization of the refunctionalization process. As a basic information-structuring device, the use of focus is ubiquitous in Spanish (and many, if not all, other languages). In this sense, it might be possible to argue that there was a higher likelihood for speakers to use the distinction between focus and background as a model for the refunctionalization of the two types of personal pronouns than other, less frequent patterns in the language.

In line with these considerations, we can hypothesize that when an obsolescent construction is refunctionalized, the linguistic pattern that serves as a model for the new function occurs in a comparable usage context (i.e., there is some semantic similarity between the refunctionalized construction and the source pattern) and has a high usage frequency.

**Hypothesis 2.** *When refunctionalizing a construction, speakers typically refer to patterns as models for the new function that (a) occur in comparable usage contexts and (b) have a high usage frequency.*

### 3. Folk Etymology as Refunctionalization

As stated in the introduction, this paper aims at giving experimental proof for Hypotheses 1 and 2. In this section, I will argue that the process of folk etymology is an adequate testing ground for the study of the influence of usage frequency on refunctionalization. Blank (2001) defines folk etymology as

a type of reanalysis. Due to their formal phonetic similarity, speakers relate two words to each other. This reanalysis always contradicts the real etymology of the reanalyzed word. (Blank 2001, p. 91, transl. MR)

Olschansky (1996) gives a more detailed definition of folk etymology:

Folk etymology is a process in which a synchronically isolated and as such unmotivated word or word constituent is attributed to a word that is phonetically similar or (partially) identical [ . . . ] in a way that is incorrect from an etymological and diachronic perspective. Consequently, the word or word constituent receives a new motivation and interpretation, and is de-isolated. (Olschansky 1996, p. 107, transl. MR)

The quote from Olschansky makes obvious that the type of reanalysis involved in refunctionalization is not syntactic but semantic reanalysis, i.e., a meaning change. To give an example, the German word *Hebamme* “midwife” derives from Old High German *HEVIANNA* “old woman who lifts a child”. Whereas the first part of the compound *HEVI-* remounts to the verb *heben* “lift,” the origin and exact meaning of the second part of the compound *-ANNA* is unknown (Kluge 2003). In later times, the morpheme *-ANNA* was replaced with the German word *Amme* “nurse,” leading to the creation of the Modern German word *Hebamme* “midwife”.

The example illustrates that folk etymology is in the first place a formal change in that the phonetic substance of the target morpheme *-ANNA* is changed in accordance with the supposed etymon (*Amme*).

Semantic factors are often irrelevant for this process as the meaning of the alleged etymon does not necessarily have to be compatible with the overall meaning of the target word, as shown by Maiden (2008, pp. 311–19). Consequently, “speakers are not seeking to ‘explain’ the meaning of a word, but to give it a familiar inner structure.” (Maiden 2008, p. 315). In this way, folk etymology does not give a *semantic motivation* to (parts of) words. However, at least to some degree, folk etymology can be characterized as a refunctionalization process. Note that in our example, at least one of the morphemes forming the original no longer has a meaning in language and is in this sense devoid of a function.<sup>1</sup> By substituting the morpheme with another morpheme, speakers arguably also confer the meaning of the alleged etymon to the resulting lexeme. As argued by Maiden (2008, p. 317), although the meaning of the morpheme *butter* is irrelevant for the meaning of the compound *butterfly*, upon hearing the compound *butterfly*, his “pre-theoretical native-speaker intuition about this compound is that the first element *is* the formative *butter*”. The experimental studies in Libben and De Almeida (2002) and Jarema (2005, pp. 47–51) cited by Maiden confirm this intuition in that speakers appear to indeed parse formatives such as *butter* in *butterfly* as the word *butter* with its specific semantics. Returning to our example of German *Hebamme*, we can argue that the reanalysis of the original morpheme –ANNA did indeed lead to a refunctionalization of HEVIANNA in that it obtained a new semantics and consequently, function. It is in this sense that folk etymology serves the function of reincorporating “orphaned words” into the lexicon (Blank 2001, p. 92).

If folk etymology can be characterized as a kind of refunctionalization process, we can use it to test the two hypotheses on the influence of usage frequency on refunctionalization established in Section 2. Regarding the question of whether or not speakers will produce a folk etymology process, we can assume that the usage frequency of the reanalyzed element plays an important role. In many folk etymology processes, the reanalyzed element is no longer used independently in the language and its meaning is no longer identifiable. In other words, the element has been affected by an obsolescence process. In line with Hypothesis 1 established in Section 2, this makes the prediction that participants are more likely to accept the false etymology if the target word has a low usage frequency.

We can also try to make predictions regarding the elements that are used as models for the reanalyzed element in folk etymologies. As suggested by the studies cited above, speakers typically select false etymons that bear a phonological resemblance to the target morpheme. A second prediction is therefore that folk etymology processes become more likely if there are elements that are formally similar to the target morpheme. However, we can also assume an influence of usage frequency in that, in line with Hypothesis 2, speakers are more likely to establish a false etymology if the alleged etymon has a high usage frequency.

#### 4. Questionnaire Study

In this section, I describe the exploratory questionnaire study<sup>2</sup> conducted in order to test Hypotheses 1–2.

##### 4.1. Materials

The questionnaire study aimed at establishing the influence of usage frequency on folk etymology processes, and consequently, refunctionalization processes, in Modern Spanish. Speakers were presented with a target word from Modern Spanish. They were then presented with an invented, false etymon for this target word and asked whether they thought it possible that the target word derived from the alleged etymon.<sup>3</sup> They were given three possible answers: “Yes,” “No,” and “I don’t know

<sup>1</sup> Indeed, instances such as this one might be the closest to Lass’ (1990) notion of *junk morphology* that can be found in a language.

<sup>2</sup> All participants were informed about the aims of the study and the anonymity of their responses, and they provided their consent.

<sup>3</sup> It would, of course, have been possible to also include factual etymologies as a control group, or even give the participants the choice between the false and the factual etymology. For the sake of simplicity of interpretation of the results, I did not consider this option in this experiment. However, I do believe that it would be viable in follow-up studies.

the word(s)”. In the case of the first answer, the participants thus accepted the folk etymology. In the case of the second answer, they rejected it. Cases in which the participants selected the third answer were eliminated from subsequent analysis.

I selected 30 Spanish target words on the basis of [Dworkin \(2012\)](#). The selection process was mostly guided by the possibility of finding a possible false etymon that was phonologically and semantically plausible for the participants of the experiment. An entire list of the materials can be found in the Appendix A (Table A1). I extracted the usage frequency (per million words) of each of the target words and the proposed etymons from the 20th century part of the *Corpus del Español* ([Davies 2002](#), over 20 million words).

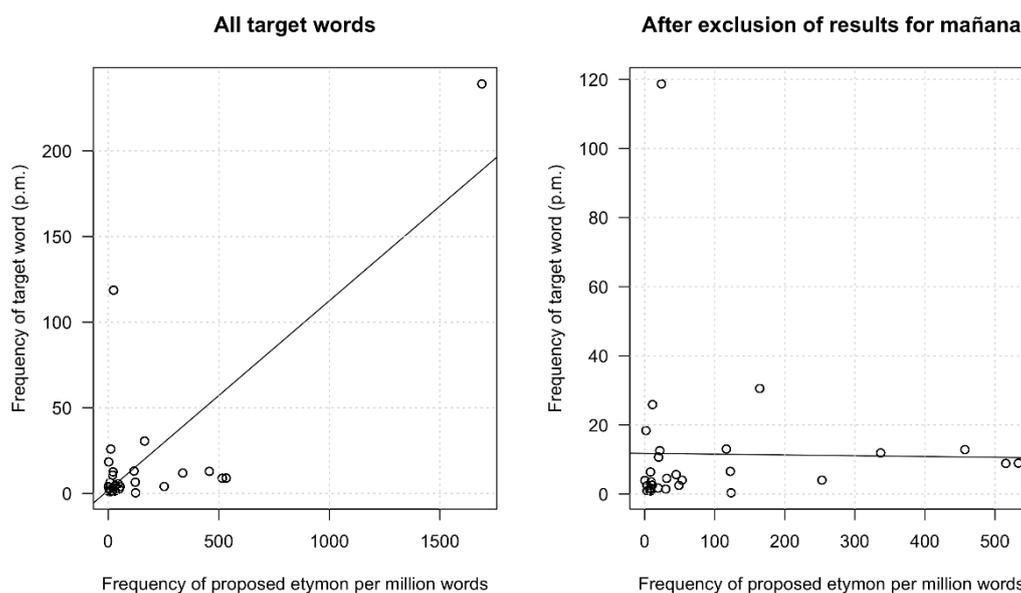
#### 4.2. Procedure

I created a questionnaire on *Google Docs* that is available in its entirety in the Appendix A (Table A2). The questionnaire was structured in four parts. After a brief introduction (Part 1), in Part 2 the participants were given a number of background questions about their sex, age, native tongue, and education. In Part 3, the participants were presented with the 30 questions about the target words. The order of the questions was randomized in order to neutralize priming effects. The questionnaire closed with a brief confirmation text (Part 4).

The questionnaire was distributed via Facebook, several classes in the Humanities Departments at the Universitat de Barcelona, as well as classes organized by the Consejo Superior de Investigaciones Científicas (CSIC) in Madrid. Sixty-seven participants took part in the questionnaire study, of which 42 were female and 25 were male. The results from 11 participants that were not native speakers of Spanish were judged unreliable and therefore eliminated from subsequent analysis, leaving a total of 56 participants.

#### 4.3. Results

When analyzing the results from the questionnaire study, I detected an unexpected and unwanted strong positive correlation between the usage frequencies of the target words and the usage frequencies of the proposed etymons ( $r = 0.79, p < 0.001$  \*\*\*). A closer look at the distribution of the two frequency measures shows that this correlation hinged on the target word *mañana* “tomorrow”, for which the false etymon *año* “year” was given. While *mañana* has by far the highest usage frequency of all target words, *año* by far has the highest usage frequency of all of proposed etymons (see Figure 1). *Mañana* can therefore be characterized as an outlier.



**Figure 1.** Correlation between the usage frequencies of the target words and the proposed etymons.

Since (a) I aimed at including both usage frequency of target words and usage frequency of proposed etymon as predictor variables in a logistic regression analysis, and (b) it is a prerequisite for logistic regression analyses that the predictor variables be independent from each other, I eliminated the results for the target word *mañana* from the analysis. This procedure also eliminated the correlation between the two predictor variables in the data (see the right plot in Figure 1). I therefore did not eliminate the results for the target word *prensa*, which also has a high usage frequency (118.7 p.m.), from the data. See Figure A1 in the Appendix A for an overview of the frequencies of acceptance for each of the target words.

Mixed-effects regression models (Baayen 2008, chp. 7; Pinheiro et al. 2018) are an adequate tool for the analysis of data from experiments such as questionnaire studies. The quality of experimental data frequently suffers from the fact that the group of experiment participants is a random sample from the global population. Each of these participants may display idiosyncrasies regarding the variable investigated in the experiment. In the case of this questionnaire study, some participants were overall more likely to answer YES to the questions about the etymology posed to them than other participants, a fact that our predictor variables, such as usage frequency, cannot explain. Mixed-effects regression models address this problem by including both *fixed effects* (the traditional predictors, i.e., repeatable factors) and *random effects* (variables such as participant, which are not repeatable). In addition to calculating the coefficients for the predictor variables, a mixed-effects regression model will calculate a random intercept for each level of the random variables, thus greatly enhancing the statistical resolution and controlling for unwanted variation due to the fact that the participants were chosen randomly and might display individual preferences.

I calculated a logistic mixed-effects regression model in R (R Development Core Team 2015) over the responses of the participants to the questions regarding the etymology of the target words (variable REFUNCTION with the outcomes “yes” or “no”). The random variable was PARTICIPANT, allowing the model to control for interpersonal variation. Table 1 summarizes the predictor variables used in the model. I decided not to include a variable measuring the education of the participants because a large majority had university education. I did however include variables referring to the length of the target words and proposed etymons (TW.LETTERS and PE.LETTERS), as well as an interaction effect between the two variables as an approximant of the importance of formal similarity between target words and proposed etymons.

**Table 1.** Summary of predictor variables.

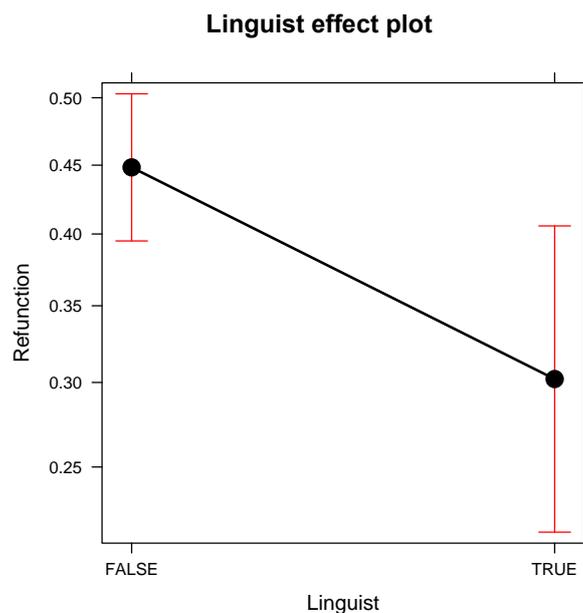
Variable Name	Short Description	Levels
AGE	Age of participant	<20, 20–30, >31
SEX	Sex of participant	f, m
LINGUIST	Whether or not the participant has university education in linguistics	yes, no
TW.LETTERS	Length of target word in letters	(numeric, z-standardized)
PE.LETTERS	Length of proposed etymon in letters	(numeric, z-standardized)
TW.FREQUENCY	Frequency of target word per million	(numeric, z-standardized)
PE.FREQUENCY	Frequency of proposed etymon word per million	(numeric, z-standardized)

In the following description of the results from the regression model, I provide plots of the effects because these are easier to interpret for readers without knowledge of regression models. The full results from the regression model can be found in the Appendix A (Table A3). The effect plots were created using the package <effects> in R.

With a  $c$  index of concordance = 0.74, the model reaches a rather moderate fit to the data that hints at the fact that important factors are missing from the analysis that explain the participants' decisions in the survey.<sup>4</sup>

#### 4.3.1. Social Indicators

Of the social indicators, only the variable LINGUIST had a significant effect on the participants' choices. AGE and SEX appeared to be irrelevant. As illustrated in the effect plot (Figure 2), participants with a knowledge of linguistics were significantly more skeptical regarding the proposed etymologies.



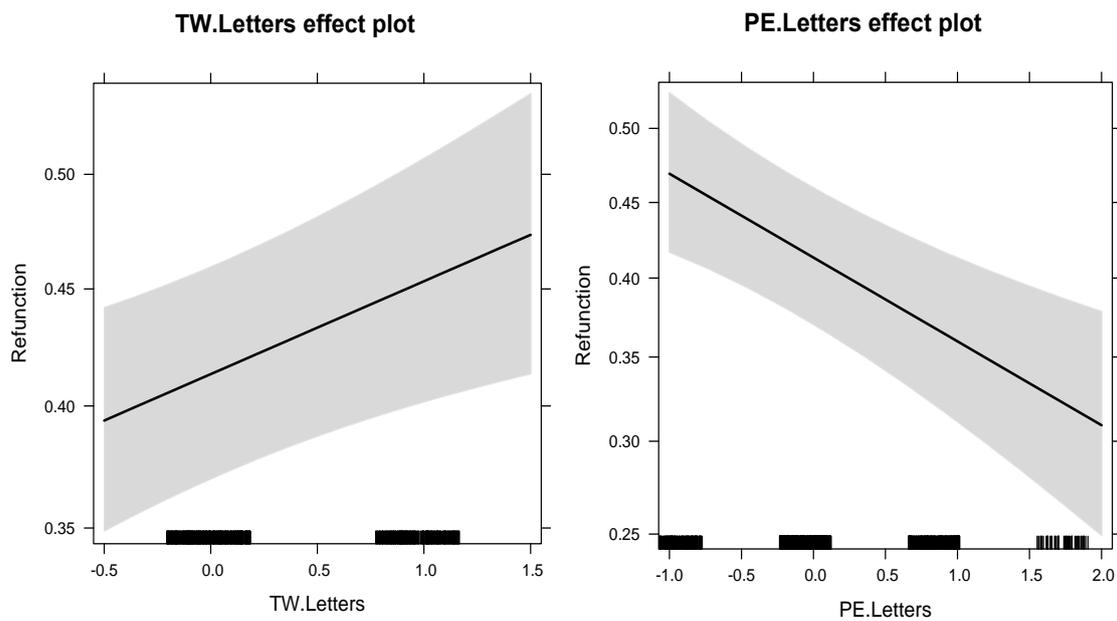
**Figure 2.** Effect plot for LINGUIST.

#### 4.3.2. Length of Target Word and Proposed Etymon

Interestingly, both the length of the target word and the proposed etymon in letters had significant and differing effects on the participants' judgment on the proposed etymologies. Whereas participants were more likely to accept the false etymology when the target word was longer (Figure 3, left plot), for longer proposed etymons, they were less likely to accept the false etymology (Figure 3, right plot).<sup>5</sup>

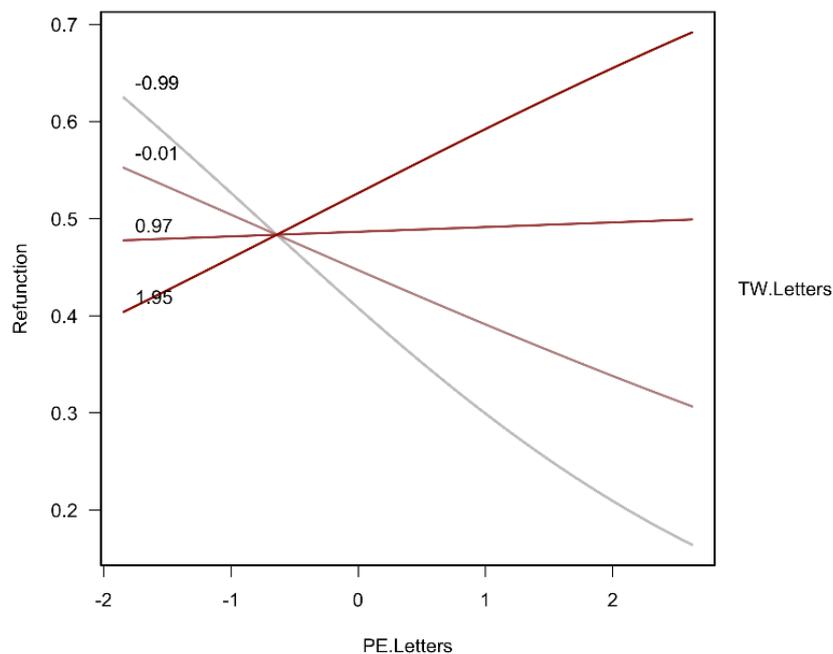
<sup>4</sup> See, for instance, Baayen (2008, p. 244), who claims that “a value above 0.8 indicates that model may have some real predictive capacity”.

<sup>5</sup> Recall that the values for all numerical variables in the analysis were z-standardized. The values on the x-axis in all of the graphs in Figures 3–5 refer to these normalized values and not the original numerical values.



**Figure 3.** Effect plot for TW.LETTERS and PE.LETTERS.

Furthermore, the analysis demonstrated the existence of an interaction effect between the two word-length indicators. Figure 4 illustrates this moderating effect of TW.LETTERS on PE.LETTERS. The finding that shorter proposed etymons are judged to be better etymons was restricted to cases in which the target word was also short (the lines marked as  $-0.99$  and  $-0.01$ ). If the target word was long (the lines marked as  $0.97$  and  $1.95$ ), the effect was reversed; participants were more likely to accept longer proposed etymons than shorter proposed etymons.



**Figure 4.** Effect plot for TW.LETTERS:PE.LETTERS (interaction effect).

### 4.3.3. Usage Frequency of Target Word and Proposed Etymon

The regression did not find a significant effect for the usage frequency of the target word on the participants' judgments on the false etymology (see Figure 5, left plot). In contrast, there was a significant effect of the usage frequency of the proposed etymon (Figure 5, right plot). Thus, participants were significantly more likely to accept the false etymology for high-frequency false etymons than for low-frequency false etymons.

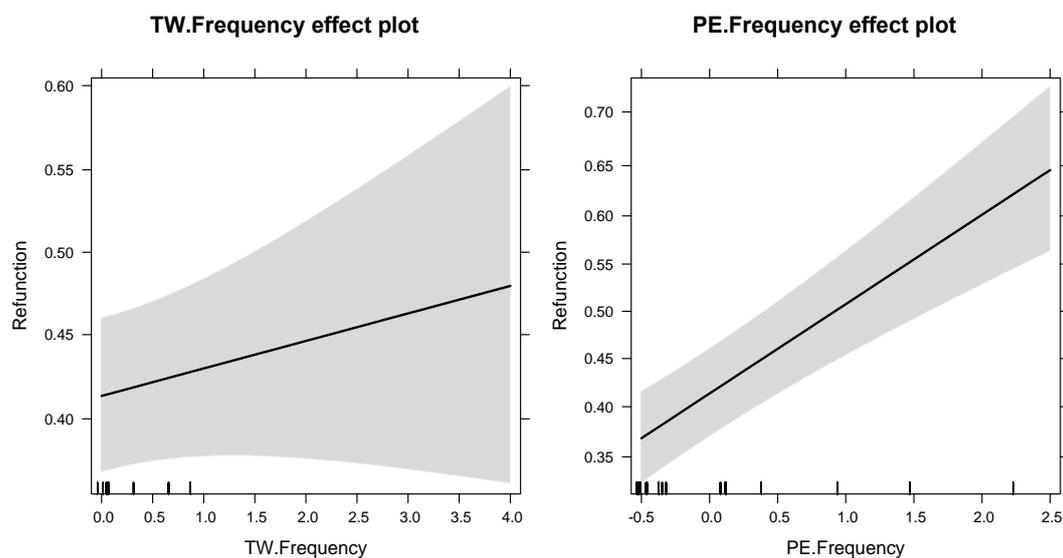


Figure 5. Effect plot for TW.FREQUENCY and PE.FREQUENCY.

## 5. Discussion

The results described in the last section confirm Hypothesis 2 in that participants were more likely to accept the false etymology when the proposed etymon was of high usage frequency than when it was of low frequency. This suggests that in processes of folk etymology, speakers will more frequently recur to high-frequency words than to low-frequency words as models. This observation is in line with usage-based approaches to language change; high-frequency words are more salient in a language than low-frequency words and therefore have a higher degree of cognitive accessibility than low-frequency words.

In contrast, Hypothesis 1—the assumption that refunctionalization and likewise, folk etymology processes, are more likely for low-frequency target forms than high-frequency target forms—was not confirmed in my analysis. This result can possibly be explained by the artificial experimental setup. Crucially, participants were only asked whether they accepted the false etymology for a target word. In real-life processes of folk etymology, speakers' motivations for folk etymology are very different. Given that the decision to produce a folk etymology process was, as it were, imposed on the participants, this might have influenced their decision processes regarding whether or not to accept the proposed etymology. In other words, the questionnaire rather measured which false etymons were judged to be better etymons than others than whether or not a folk etymology process should take place.

Additionally, the regression analysis suggested that folk etymology processes were more likely for shorter proposed etymon words and longer target words. It is well known that shorter words are typically more frequent and therefore have a higher degree of cognitive accessibility (Bybee 2010, pp. 20–21). Consequently, one could argue that the participants were more likely to accept shorter proposed etymons because of their higher degree of cognitive accessibility. However, given that the results for target words point in the opposite direction, it appears to me that it is incorrect to invoke cognitive accessibility as the relevant parameter. Rather, it appears that the participants have

an understanding (either intuitive or due to their training as linguists) that due to derivation or composition processes, words that are historically derived from certain source words tend to be longer than these source words.

The interaction effect between the length of the target word and the length of the proposed etymon analysis appears to reveal that structural similarity also had an influence on the participants' decisions to accept a proposed etymology. Participants were more likely to accept a proposed etymology if (a) both the target word and the proposed etymon were short or (b) both the target word and the proposed etymon were long.

On a final note, the regression model was not able to explain much of the variance in the results, as indicated by the relatively low value of the *c* index of concordance. This lack of statistical resolution suggests that the analysis was lacking crucial parameters. In particular, there were no parameters measuring the degree of formal and semantic similarity between the target words and the proposed etymons. It seems likely that by including psycholinguistics measurements of the degree of similarity between these items it would be possible to greatly enhance the statistical resolution of the analysis and uncover more effects on the participants' behavior in the questionnaire survey.

## 6. Conclusions

In this paper, I hypothesized that usage frequency plays an important role in refunctionalization processes in that refunctionalization is more likely for low-frequency constructions than high-frequency constructions (Hypothesis 1), and that high-frequency constructions are better models for refunctionalization processes than low-frequency constructions (Hypothesis 2). I proposed that folk etymology can be regarded as a type of refunctionalization process and that it can therefore be used in order to empirically test these assumptions. The results from the questionnaire study on folk etymology supported some of the assumptions. In particular, participants were more likely to accept false etymons for the target words if the proposed etymons had a high usage frequency, confirming Hypothesis 2. The experiment was however not able to confirm Hypothesis 1. I also found an effect of the length of the target word, as well as the length of the proposed etymon. It was also found that participants with a background in linguistics fared significantly better in the survey than participants without knowledge of linguistics.

It must be noted that these findings are preliminary and should therefore be taken with caution. In particular, the low degree of variance explained by the regression analysis indicates that the study was missing crucial parameters that may relativize some of the results from the analysis. Further studies on the reasons for folk etymology processes should consider including psycholinguistic measurements of the degree of formal and semantic similarity between target words and proposed etymons. These measurements might also provide the experimenter with a more principled way of selecting the stimuli. Lastly, there might be experiment types that are better suited than questionnaire studies for such an analysis. One could, for instance, imagine an experiment in which the reaction of participants to false etymologies is measured using reading times or even fMRI. Longer reaction times should then indicate more problems with the acceptance of the false etymology. Such response types would serve to diminish the degree to which participants actively contemplate the likelihood of the proposed etymology, and consequently increase the reliability of the experimental results.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A

Table A1. Target words and proposed etymons.

Target Word	Etymology (Dworkin 2012; Real Academia Española 2014)	Proposed Etymon
<i>Acechar</i> "to stalk"	lat. * ASSECTARE "attend to" (Dworkin 2012, p. 102)	<i>echar</i> "to throw"
<i>achacar</i> "to blame"	arab. TASAKKA "blame" (Dworkin 2012, p. 102)	<i>atacar</i> "to attack"
<i>ademán</i> "gesture"	arab. AD-DAMAN "legal guarantee" (Dworkin 2012, p. 46)	<i>Mano</i> "hand"
<i>adorar</i> "to adore"	lat. AD-ORARE "towards-pray" (Dworkin 2012, p. 161)	<i>oro</i> "gold"
<i>albañil</i> "bricklayer"	arab. AL-BANNA "the construer" (Dworkin 2012, p. 112)	<i>Baño</i> "bathroom"
<i>almirante</i> "admiral"	arab. AMÎR "commander" (Dworkin 2012, p. 90)	<i>mirar</i> "to look"
<i>alojar</i> "to lodge"	sp. LONJA "portico, porch" (Dworkin 2012, p. 146)	<i>lugar</i> "place"
<i>amainar</i> "to weaken, flag"	goth. AF-MAGINON "lower the sails" (Dworkin 2012, p. 72)	<i>marginal</i> "marginal"
<i>apagar</i> "to turn off"	lat. PACARE "pacify, quiet" (Dworkin 2012, p. 55)	<i>pagar</i> "to pay"
<i>azafata</i> "stewardess"	arab. AL-SAFÁT "basket" (Dworkin 2012, p. 109)	<i>zarpar</i> "to set sails"
<i>bagaje</i> "baggage"	fr. BAGAGE "baggage" (Dworkin 2012, p. 127)	<i>vagar</i> "to wander"
<i>barrer</i> "to sweep"	lat. VERRERE "sweep" (Dworkin 2012, p. 46)	<i>barro</i> "mud"
<i>bigote</i> "mustache"	ger. BEI GOTT "with god" or fr. BIGOT (Dworkin 2012, pp. 77–78)	<i>gota</i> "drop"
<i>bisoño</i> "greenhorn"	it. BISOGNO "need" (Dworkin 2012, pp. 151–52)	<i>sueño</i> "dream"
<i>bochorno</i> "extreme heat"	lat. VULTURNUS "south wind" (Dworkin 2012, p. 49)	<i>horno</i> "oven"
<i>borrasca</i> "storm (at sea)"	probably it. BURRASCA "storm" (Dworkin 2012, p. 145)	<i>rascar</i> "to scratch"
<i>bostezar</i> "to yawn"	lat. OSCITARE "open the mouth wide" (Dworkin 2012, p. 52)	<i>voz</i> "voice"
<i>carabela</i> "type of ship"	ptg. CARAVELA "type of ship" < greek κάραβος "light boat" (Real Academia Española 2014)	<i>vela</i> "to sail"
<i>cerveza</i> "beer"	celtic CERVISIA "beer" (Dworkin 2012, p. 28)	<i>hervir</i> "to boil"
<i>debilidad</i> "weakness"	fr. DEBILITÉ "weakness" (Dworkin 2012, p. 132) < PIE *BEL "power, strength"	<i>bilis</i> "bile"
<i>despejado</i> "cloudless"	ptg. DESPEJAR "pour" (Dworkin 2012, p. 188)	<i>espejo</i> "mirror"
<i>enojar</i> "anger"	o.provenz. ENOJAR "anger" (Dworkin 2012, p. 125)	<i>ojo</i> "eye"
<i>escopeta</i> "shotgun"	it. SCHIPETTO "firearm" (Dworkin 2012, p. 152)	<i>escupir</i> "to spit"
<i>esgrimir</i> "to wield a weapon"	o.provenz. ESGREMIR "wield a weapon" (Dworkin 2012, p. 125)	<i>grima</i> "chills"
<i>eslabón</i> "link (of a chain)"	goth. SNÔBÔ "link" (Dworkin 2012, p. 72)	<i>eslavo</i> "slave"
<i>espuela</i> "spur"	goth. SPAÚRA "spur" (Dworkin 2012, p. 71)	<i>esposo</i> "husband"
<i>iguana</i> "type of saurian"	taíno IGUANA "type of saurian" (Dworkin 2012, p. 200)	<i>guante</i> "glove"
<i>jarabe</i> "syrup"	arab. SHARB "syrup" (Dworkin 2012, p. 84)	<i>jarro</i> "jug"
<i>mañana</i> "morning"	v.lat. *MANEANA "morning" (Dworkin 2012, p. 56)	<i>año</i> "year"
<i>prensa</i> "press"	cat. PREMIER "press, squeeze" (Dworkin 2012, p. 194)	<i>prender</i> "to take"

Table A2. Questionnaire design.

Original Text	English Translation
<p><b>1. Introduction</b></p> <p>¡Hola! Muchas gracias por cooperar en nuestro EtimoTest. Pretendemos examinar el conocimiento que tienen de la etimología—el origen de las palabras—los hablantes nativos del español.</p> <p>A menudo, las nuevas palabras se basan en otras palabras ya existentes en la lengua. Por ejemplo, el adjetivo barato deriva del verbo baratar (“trocar, comprar a bajo precio”). La palabra escarnimiento (“desengaño”) deriva del antiguo verbo escanir (“hacer burla de alguien”). A veces tenemos una intuición sobre la palabra que fue la base para una nueva palabra; otras veces, en cambio, no la tenemos.</p> <p>Para examinar tu nivel de conocimiento de la etimología de las palabras, te presentaremos 30 palabras españolas. Para cada una de ellas, te ofrecemos una palabra base como solución. En algunos casos, esta palabra es la palabra base correcta; en otros casos, no. Te preguntaremos si es posible que la palabra derive de la palabra base, y te daremos tres opciones: SÍ, NO, y NO CONOZCO LA(S) PALABRA(S). Por favor, selecciona esta última opción solo si no conoces una de las palabras. En total, el EtimoTest dura alrededor de 10 minutos.</p> <p>Un aviso importante: por favor rellena el cuestionario sin utilizar recursos como diccionarios, google, etc. Así invalidarías los resultados de tu EtimoTest.</p> <p>Antes de empezar, tenemos que hacerte unas breves preguntas sobre tu persona. Te garantizamos que tus respuestas van a ser tratadas con la máxima discreción.</p>	<p><b>1. Introduction</b></p> <p>Hi! Thank you very much for participating in our EtimoTest. We want to investigate the knowledge of etymology—the origin of words—of native speakers of Spanish.</p> <p>Frequently, new words are based on other words that already exist in a language. For instance, the adjective barato “cheap” derives from the verb baratar “to bargain, buy at a cheap price”. The words escarnimiento “punishment” derives from the old verb escanir “make fun of somebody”. Sometimes we have an intuition about which word served as a basis for the new word, sometimes we do not.</p> <p>In order to examine your level of knowledge of the etymology of words, we will present you with 30 Spanish words. For each one of these, we offer you an origin word as a solution. In some cases, this word is the correct origin word, in some cases it is not. We will ask you if it is possible that the word derives from the origin word, and we will give you three options: YES, NO, and I DO NOT KNOW THE WORD(S). Please select this last option only if you do not know one of the words. The EtimoTest will last around 10 min.</p> <p>One important point: please fill out the questionnaire without using dictionaries, Google, etc. In doing so, you would invalidate the results of your EtimoTest.</p> <p>Before we begin, we have to ask you a few questions about yourself. We guarantee you that your answers will be treated with maximal discretion.</p>
<p><b>2. Background questions</b></p> <p>2.1 ¿Cuál es tu sexo? Femenino Masculino</p> <p>2.2 ¿Cuál es tu edad? Menos de 20 años Entre 20 y 30 años Más de 30 años</p> <p>2.3 ¿Eres hablante nativo del español? Sí No</p> <p>2.4 ¿Cuál es tu nivel de educación? Estudios Primarios Educación Secundaria Obligatoria Bachillerato Educación secundaria post obligatoria Estudios universitarios (Grado/Máster/Posgrado/Doctorado)</p> <p>2.5 Sí estás realizando/has realizado estudios universitarios, ¿cuál era la asignatura/las asignaturas?</p>	<p><b>2. Background questions</b></p> <p>2.1 What is your sex? Feminine Masculine</p> <p>2.2 What is your age? Less than 20 years Between 20 and 30 years More than 30 years</p> <p>2.3 Are you a native speaker of Spanish? Yes No</p> <p>2.4 What is your level of education? Estudios Primarios Educación Secundaria Obligatoria Bachillerato Educación secundaria post obligatoria Estudios universitarios (Grado/Máster/Posgrado/Doctorado)</p> <p>2.5 If you are studying/have studied at a university, what was the study subject?</p>

Table A2. Cont.

Original Text	English Translation
<b>3. Test</b>	<b>3. Test</b>
¿Es posible que la palabra [TARGET] derive de la palabra [PROPOSED ETYMON]?	Is it possible that the word [TARGET] derives from the word [PROPOSED ETYMON]?
Sí	Yes
No	No
No conozco la(s) palabra(s)	I do not know the word(s)
[30 questions in total, randomized order]	[30 questions in total, randomized order]
<b>4. Confirmation text</b>	<b>4. Confirmation text</b>
Esta ha sido la última pregunta. ¡Muchas gracias por tu participación en el EtimoTest!	This has been the last question. Thank you very much for your participation in the EtimoTest!

Table A3. Full results from the logistic regression model.

Variable	Level	Betas	Odds	SE	ZVAL	p
(Intercept)		0.00	1.00	0.30	0.10	0.994
AGE	<20			<i>Reference level</i>		
	21–30	−0.18	0.84	0.31	−0.58	0.562
	>31	−0.44	0.64	0.32	−1.39	0.164
SEX	Feminine			<i>Reference level</i>		
	Masculine	0.15	1.16	0.20	0.74	0.458
LINGUIST	FALSE			<i>Reference level</i>		
	TRUE	−0.63	0.53	0.27	−2.33	0.020
TW.LETTERS		0.16	1.18	0.06	2.86	0.004
PE.LETTERS		−0.23	0.80	0.06	3.77	0.000
TW.FREQUENCY		0.07	1.07	0.06	1.17	0.243
PE.FREQUENCY		0.38	1.46	0.06	6.35	0.000
TW.LETTERS: PE.LETTERS		0.25	1.29	0.07	3.77	0.000

Number of observations = 1629  
C index of concordance = 0.74  
Somers' dxy = 0.47  
AIC = 2082.2  
BIC = 2141.5

MODEL EVALUATION

p values: \* = <0.05, \*\* = <0.01, \*\*\* = <0.001

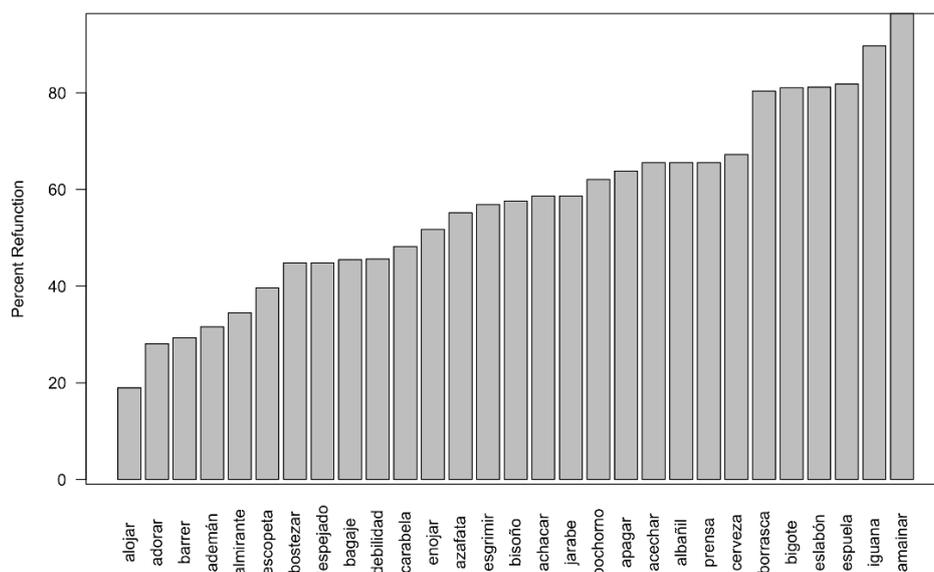


Figure A1. Proportion of accepted proposed etymologies by target words.

## References

- Baayen, Harald. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Blank, Andreas. 2001. *Einführung in die lexikalische Semantik für Romanisten*. Tübingen: Niemeyer.
- Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82: 711–33. [CrossRef]
- Bybee, Joan L. 2010. *Language, Usage, and Cognition*. Cambridge: Cambridge University Press.
- Davies, Mark. 2002. Corpus del Español (100 Million Words, 1200s–1900s). Available online: <http://www.corpusdelespanol.org> (accessed on 22 December 2015).
- De Smet, Hendrik. 2012. The course of actualization. *Language* 88: 601–33. [CrossRef]
- Dworkin, Steven N. 2012. *A History of the Spanish Lexicon: A Linguistic Perspective*. Oxford: Oxford University Press.
- Jarema, Gonia. 2005. Compound representation and processing: A cross-language perspective. In *Representation and Processing of Compound Words*. Edited by Gary Libben and Gonia Jarema. Oxford: Oxford University Press, pp. 45–70.
- Kluge, Friedrich. 2003. *Etymologisches Wörterbuch der deutschen Sprache*. Berlin: De Gruyter.
- Lass, Roger. 1990. How to do things with junk: Exaptation in language evolution. *Journal of Linguistics* 1: 79–102. [CrossRef]
- Libben, Gary, and Roberto G. De Almeida. 2002. Is there a morphological parser? In *Morphology 2000*. Edited by Sabrina Bendjaballah, Wolfgang U. Dressler, Oskar E. Pfeiffer and Maria D. Voeikova. Amsterdam and Philadelphia: John Benjamins, pp. 213–25.
- Mackenzie, Ian. 2006. *Unaccusative Verbs in Romance Languages*. Basingstoke: Palgrave Macmillan.
- Maiden, Martin. 2008. Lexical nonsense and morphological sense: On the real importance of 'folk etymology' and related phenomena for historical linguists. In *Grammatical Change and Linguistic Theory: The Rosendal Papers*. Edited by Thórhallur Eythórsson. Amsterdam and Philadelphia: John Benjamins, pp. 307–28.
- Narrog, Heiko. 2016. Exaptation in Japanese and beyond. In *Exaptation and Language Change*. Edited by Muriel Norde and Freek Van de Velde. Amsterdam and Philadelphia: John Benjamins, pp. 93–120.
- Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.
- Norde, Muriel, and Freek Van de Velde, eds. 2016. Exaptation. Taking stock of a controversial notion in linguistics. In *Exaptation and Language Change*. Amsterdam and Philadelphia: John Benjamins, pp. 1–35.
- Olschansky, Heike. 1996. *Volksetymologie*. Berlin: De Gruyter.
- Pinheiro, Jose, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and the R Development Core Team. 2018. *nlme: Linear and Nonlinear Mixed Effects Models*, R Package Version 3.1-137; Available online: <https://CRAN.R-project.org/package=nlme> (accessed on 23 October 2018).
- Poplack, Shana. 2001. Variability, frequency, and productivity in the irrealis domain of French. In *Frequency and the Emergence of Linguistic Structure*. Edited by Joan L. Bybee and Paul J. Hopper. Amsterdam and Philadelphia: John Benjamins, pp. 405–28.
- Poplack, Shana, Allison Lealess, and Nathalie Dion. 2013. The evolving grammar of the French subjunctive. *Probus* 25: 139–95.
- R Development Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Available online: <http://www.R-project.org> (accessed on 26 December 2015).
- Real Academia Española. 2010. *Nueva gramática de la lengua española. Manual*. Madrid: Escasa Libros.
- Real Academia Española. 2014. *Diccionario de la lengua española*. Madrid: Real Academia Española.
- Rodríguez Molina, Javier. 2006. *Ser + participio en español antiguo: Perífrasis resultativa, no tiempo compuesto*. In *Actas del VI Congreso Internacional de Historia de la Lengua Española*. Edited by José J. Bustos Tovar and José L. Girón Alchonchel. Madrid: Arco Libros, vol. II, pp. 1059–72.
- Rosemeyer, Malte. 2014. *Auxiliary Selection in Spanish. Gradience, Gradualness, and Conservation*. Amsterdam and Philadelphia: John Benjamins.
- Smith, John C. 2006. How to do things without junk. The refunctionalization of a pronominal subsystem between Latin and Romance. In *New Perspectives on Romance Linguistics: Vol. II: Phonetics, Phonology and Dialectology. Selected Papers from the 35th Linguistic Symposium on Romance Languages (LSRL), Austin, Texas, February 2005*. Edited by J. Montreuil. Amsterdam and Philadelphia: John Benjamins, pp. 183–205.

Willis, David. 2010. Degrammaticalization and obsolescent morphology: Evidence from Slavonic. In *Grammaticalization: Current Views and Issues*. Edited by Ekatherini Stathi, Elke Gehweiler and Ekkehard König. Amsterdam and Philadelphia: John Benjamins, pp. 151–78.



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).