Phonological properties of word classes and directionality in conversion[1]

Arne Lohmann
Heinrich-Heine-Universität Düsseldorf
arne.lohmann@hhu.de

To appear in *Word Structure* 10(2)
Pre-publication version

December 1, 2016

Abstract

In the study of the word-formation process of conversion, one particularly difficult task is to determine the directionality of the process, that is, to decide which word represents the base and which the derived word. One possibility to inform this decision that has received only limited attention is to capitalize on word-class-specific phonological properties. This paper empirically investigates this option for English noun-verb conversion by building on recent findings on phonological differences between these two word classes. A large-scale study of phonological properties is carried out on CELEX data, employing the quantitative techniques of conditional inference trees and random forests. An important result of this analysis is that the accuracy of phonological cues varies widely across different subsamples in the data. Essentially this means that phonological cues can be used as a criterion to determine the directionality of words that are at least two syllables in length. When restricted to this part of the lexicon, phonological properties represent a fairly accurate indicator of source word class and are therefore a useful addition to the linguist's toolkit for determining directionality in conversion. Based on this result, the paper also discusses the relations of phonological properties to other criteria commonly employed to determine directionality.

1 Introduction

This paper empirically explores to what extent phonological characteristics may serve to indicate directionality in the word-formation process of conversion (or zero-derivation) between nouns and verbs in English. The starting point of this approach is the recent observation that these word classes differ in an unexpectedly large number of phonological dimensions. A well-known difference is that English disyllabic nouns and verbs differ with regard to position of stress, with nouns typically stressed on the initial and verbs on the final syllable, but a host of other systematic, albeit less conspicuous contrasts have been uncovered over the past years (see Kelly 1996, Monaghan & Christiansen 2008, Sandoval et al. 2012 for overviews).

In conversion, a word that instantiates a different syntactic category is derived from a source word, usually without overt marking. The result of this process are word pairs of two syntactic categories most of which are phonologically identical, for example, *to escape* (V) vs. *escape* (N).[2] One of the major issues in conversion research is to determine the directionality of the process, for example, deciding whether a particular word pair is an example of Noun-to-Verb (NtoV), or Verb-to-Noun (VtoN) conversion. This project tests to what extent it is possible to determine the source word of this process via word class-specific phonological characteristics, with the goal of obtaining a further criterion for the linguist to decide on directionality.

Before addressing this question, it needs to be pointed out that the issue of directionality is dependent on the theory of the lexicon one adopts. Of particular relevance is the question whether entries in the lexicon are marked for syntactic category. For example, in 'Distributed Morphology' (Halle & Marantz 1993, Marantz 1997), lexical items are stored as 'roots' without specification of their categorical status (see Barner & Bale 2002 for a similar view). Only through its use in syntax does the category of a word becomes specified. Under this view, there is no word-formation process of conversion, as roots are in principle free to occur in different syntactic contexts. Hence there is also no directionality, as it is not possible to distinguish a base from a derived word, but merely different contextual uses of the same root (see Farrell 2001 for a compatible view from a cognitive linguistic perspective).

However, arguments against this 'underspecification' view have been put forward, pointing out that it is at odds with a number of empirical findings. Don (2004, 2005) provides morphological and phonological evidence for the assumption that a basic and a derived word can be distinguished in English and Dutch noun-verb pairs. The general thrust of Don's (2004, 2005) arguments is that there are certain asymmetries between different groups in the data that can be explained by their instantiating different directions of conversion. One of these arguments (see Don 2004: 938) is based on the differential stress patterns of English disyllabic noun-verb pairs, as described in Marchand (1969: 378-379) and Kiparsky (1982). Some of these pairs, such as *torment*, exhibit stress shift, where the verb is stressed on the second syllable, and the noun on the first syllable. Other pairs do not show stress shift; for example, the noun *PATtern* and the verb *PATtern* are both stressed on the initial syllable. This difference is addressed in Kiparsky's (1982) model of lexical phonology, which states that torment is lexically a verb and therefore affected by the stress rules for verbs (which means stress on the second syllable), while pattern is lexically a noun and is stressed accordingly. By postulating that VToN conversion operates before the stress rules, while NToV conversion applies after them, Kiparsky's (1982) model is able to explain why some pairs are stress-shifting while others are not. The crucial point for the present discussion is that this argument relies on a distinction between two different directions of the conversion process. It is difficult to see how a theory in which lexical entries are not specified for syntactic category could account for the fact that some noun-verb pairs shift stress, while others do not. While I do not want to repeat the discussion on directionality in detail here, the current paper sides with the view expressed by Don (2004, 2005), namely that conversion is regarded as a word-formation process with a specifiable input and output, and is therefore clearly directional. This is also the assumption held in most reference works on word-formation (e.g., Plag 2003, Bauer, Lieber & Plag 2013).

If viewed as such, the determination of the particular direction of conversion becomes highly relevant for a deeper understanding of this word-formation process. More specifically, only the determination of directionality allows for a quantitative analysis of which process, NtoV, or VtoN, is more frequent and potentially more productive. Furthermore, the different semantic relationships established through conversion can only be analyzed if the input and the output of the process are identified (see for example the semantic patterns of denominal verbs identified by Marchand 1963 and Clark & Clark 1979, and the discussions of the semantic range of conversion for converted nouns and verbs, respectively, in Bauer, Lieber & Plag 2013). Finally, the discussion of a possible competition between conversion and other derivational processes requires knowledge of base and derived word (see the discussion in Plag 1999).

A number of criteria have been suggested to guide decisions on directionality, all of which ultimately go back to the foundational works of Marchand (1963, 1964) on the topic (see Bram 2011 for an overview). These criteria have in common that they test whether one of the two words of a noun-verb pair shows signs of being derived from the other. In that sense, these criteria can be understood as not only determining the particular direction of conversion, but also providing evidence of the per se existence of a directional process. One approach is to apply what has been termed the historical criterion, that is, to investigate which word was used first, usually by relying on dates of first attestation of the relevant words in etymological dictionaries (e.g., Jespersen 1946, Marchand 1963, see also an overview in Bram 2011: 78-84). Other strategies do not address the diachronic development, but the synchronic relation between the two words' semantics. In identifying different semantic patterns that are established through conversion, Marchand (1964: 12) finds that either the noun or the verb meaning may be dependent 'on the content of the other pair member', which is thus to be considered the source word (see also Marchand 1963: 186). This possible dependence is discussed by Kiparsky (1997), who points out that a dependence of the verb on the noun meaning can be observed in examples such as *tape* (N/V), because the action of taping requires the object of tape. The former example can, therefore, be considered to be a denominal verb (Kiparsky 1997: 17). Examples of the reverse direction of semantic dependence have also been discussed in the literature. Marchand (1963: 186) mentions *look*, *ride*, and *walk*, whose noun forms are best described as denoting 'the act of V-ing', which thus indicates a direction of derivation from verb to noun (see also Don 2005: 100).[3]

A further semantic criterion, termed 'semantic range' (Marchand 1963: 186, Bram 2011: 128), or 'semantic complexity' (Plag 2003: 109), states that the range of meanings of the derived word is smaller compared to the source word. This may be explained by only one or a few senses of the source word being mapped to the new word during the process of conversion (see Plank 2010). The word pair *bottle* (N) vs. *bottle* (V) may serve as an example, as the noun has more listed meanings/senses in dictionaries than the verb, so that the latter can be assumed to be the derived word (see Bram 2011: 129-130). A further variable, related to semantic range, is 'a wider range of [...] usage' (Marchand 1963: 182) of one of the two words, which translates to a greater frequency of occurrence as evidenced in corpora (see Plag 2003: 111, Bram 2011). The logic underlying this criterion is that the derived word, which has a narrower range of meanings, should be used less frequently than the source word with a broader semantic range, capitalizing on the correlation between polysemy and frequency first observed by Zipf (1945).

The four criteria described above are the most commonly mentioned ones in the literature (cf. Bram 2011). Their use has been occasionally criticized, mostly because their application and operationalization are difficult (see Sanders 1988, Stekauer 1996, Bram 2011 for discussions of the value of the individual criteria). For example, relying on dates from etymological dictionaries (mostly the Oxford English Dictionary) when applying the historical criterion may not provide an accurate picture of the actual diachronic development, because written sources may mask differences in usage in speech (Bram 2011: 132). With regard to the frequency criterion, Stekauer (1996: 129) objects that confounding factors such as region, topic and idiolect may complicate its application. Testing this criterion on corpus data, Bram (2011) finds that words of lesser frequency are not always found in the data sources available (for details see the discussion in Bram 2011: 314-322). When applying the criterion of semantic dependence, it has been found difficult to decide whether there is an actual dependence between two words (see Sanders 1988: 173-4, Stekauer 1996: 127-129, Kiparsky 1997). With the criterion of semantic range, it is not clear how to determine the number of senses of the individual words. Bram (2011) does so by using dictionary definitions, but finds that these frequently choose a 'noun-based definition strategy' (Bram 2011: 297), which leads to an inflation of noun-to-verb results.

These issues do not disqualify the use of these criteria per se, but they point to certain problems in their application, due to which it is not always easy to answer the question of directionality for every individual data point. For this reason, the determination of directionality has been described as a problematic task (see Plag 2003, Balteiro 2007, Bram 2011). Some of these problems may be overcome by applying more refined methods of operationalization, or simply by more sophisticated resources becoming available to the linguist, such as larger corpora. A further opportunity for improvement would be to obtain an additional criterion that is less affected by operationalization problems. It is this approach that is pursued in the current paper, by empirically exploring phonological properties as an additional resource for deciding on directionality.

The reasoning underlying this approach is that if there are systematic differences between nouns and verbs regarding their phonological make-up, cases of NtoV conversion should still be characterized by their source 'nouny' phonology, while cases of VtoN conversion may be identified through phonological characteristics typical of verbs. For Dutch, the effectiveness of such an approach has been demonstrated by Don (2004: 943-44), who shows that an analysis of syllable structure allows for the identification of NtoV cases. For English, the only phonological property that has been discussed in this regard is the stress pattern of disyllabic words. As mentioned above, Kiparsky (1982, 1997), and Marchand (1969) before him, propose that stress-shifting pairs indicate a direction of conversion from verb to noun. Yet, such pairs make up only a minority of conversion cases. Kiparsky and Marchand suggest, however, that position of stress may indicate directionality also in the larger group of cases not characterized by stress shift, as the stress pattern of the source word is retained in the derived word. Thus trochaic stress would indicate derivation from noun to verb, as in *PAttern* (N/V), while iambic stress would indicate VToN, as in *comMAND* (N/V) (see Marchand 1969: 378-379, also Balteiro 2007: 166).[4]

Using the property of stress position in this sense seems to be a promising starting point; however, its accuracy in indicating the source word has not been empirically quantified. Moreover, recent research suggests that many more phonological characteristics distinguish nouns from verbs, whose applicability to inform decisions about directionality has not been explored yet (see e.g., Monaghan & Christiansen 2008).[5] This paper, therefore, aims to empirically test all phonological properties discussed in the relevant literature as to their suitability to determine the source word in English noun-verb conversion.

In order to be able do so, the phonological differences between the two word classes need to be known, of course. While considerable research has been conducted on the phonological characteristics of nouns and verbs in recent years, most of the relevant studies are confined to child-directed speech (e.g., Monaghan et al. 2005, Monaghan et al. 2007). Thus, it is not clear whether differences identified in these studies are a characteristic of language in general, or merely a feature of that particular type of language use. Therefore, I will, as a first step, test suggested phonological properties on the entire noun-verb lexicon as represented in CELEX, a large electronic dictionary of the English language.[6] This analysis is an important end in itself, as it will reveal whether the phonological contrasts between nouns and verbs discussed in previous research are truly a characteristic of the English lexicon. An additional aim of the paper is a methodological one, namely to demonstrate the suitability of the technique of conditional inference trees and random forests in testing phonological properties of word class. Conditional inference trees create subsamples based on the predictor variables in an iterative fashion. As I will show below, this feature renders this method particularly well suited for the task at hand, because many phonological differences between nouns and verbs hold for subsamples of the lexicon only. The remainder of the article is structured as follows. In section 2, I introduce the data sources used and explain the coding process of the different phonological variables tested. Section 3 reports the empirical analysis of the data and, based on these results, suggests different ways to use phonological cues to determine directionality in conversion. In section 4, I discuss the validity of phonological criteria in the wider context of how to determine directionality in conversion and by comparing them to other criteria typically employed to carry out that task. Section 5 concludes the paper.

2 Data and Variables

2.1 Data sources

This paper concentrates on those cases of noun-verb conversion for which the form of both words is truly identical, e.g., (1) and (2) below, thereby excluding cases of stress shift, as in (3), and (or) changes on the segmental level, as in (4) (stress shift and vowel change).

(1) profit (N) vs. profit (V)
(2) milk (N) vs. milk (V)
(3) DIgest (N) vs. diGEST (V)
(4) REcord (N) vs. reCORD (V)

The sample of noun-verb conversion cases employed for the empirical analysis is the one provided by Bram (2011), which, to my knowledge, is the largest such sample available. It contains 1,880 cases of noun-verb conversion compiled from lists found in the relevant literature and through a large-scale search of corpus and dictionary data (for details see Bram 2011: 116-121). As mentioned above, before testing phonological cues on cases of conversion, it is first necessary to explore the phonological differences between nouns and verbs in the English lexicon. For this analysis, I used the CELEX lemma database (Baayen et al. 2001), which contains the uninflected stem forms of words, and extracted all entries that were marked as either noun or verb.[7] From this list, I excluded all lemmas that had both a noun and a verb entry and, furthermore, weeded out all conversion cases contained in the Bram (2011) sample, so as to obtain only the noun-verb share of the lexicon that is unambiguous with regard to word class. I also excluded entries that are written as two words, as their status as a single lexical entry is debatable. This resulted in the exclusion of phrasal verbs, such as *to give up*, and compounds, such as *Bronze Age*.

2.2 Variables and coding of data

Based on an analysis of the recent literature on phonological characteristics of nouns and verbs (in particular see the list in Monaghan et al. 2005), I compiled a list of variables to be tested. This list is provided below, along with a description of how the individual variables were coded. Both the CELEX sample and the conversion sample were coded for these variables.

Variable 1 Stress pattern: It is well known that disyllabic nouns and verbs differ with regard to the position of stress: Nouns are usually stressed on the first syllable, while in verbs, the second syllable tends to be stressed (Bock & Kelly 1988, Sereno & Jongman 1990, Kelly 1996). Berg (2000) finds that also with trisyllabic words differences in stress pattern persist between the two word classes, with nouns exhibiting initial stress more often. I coded this variable for words longer than one syllable and used the following coding scheme: main stress on the initial syllable (1), main stress not on the initial syllable (-1).

Variable 2 Word Length: In previous research, it has been shown that, in English, nouns are longer than verbs in terms of the number of syllables (Cassidy & Kelly 1991, 2001, Berg 2000). I therefore coded length in number of syllables for all data points.

Variable 3 Syllabic Complexity: Monaghan et al. (2007) find that verbs in child-directed speech contain more complex syllables than nouns (see also Nazzi & Houston 2006). This difference will be tested here by coding for the average number of phonemes per syllable for each word (cf. Durieux & Gillis 2001).

Variable 4 Word Onset Complexity: Monaghan et al. (2005) find that verbs in child-directed speech are characterized by more complex word onsets than nouns. I coded the number of consonants in the word onset.

Variable 5 Ratio of reduced vowels: Nouns have been found to contain a higher ratio of reduced vowels than verbs in child-directed speech (Monaghan et al. 2007). This possible difference is tested here by coding the ratio of syllables whose nucleus was filled either by [ə] or a syllabic consonant.

Variable 6 Vowel Backness of the tonic syllable: Another phonological difference reported to exist between nouns and verbs is the vowel quality of the tonic syllable. Sereno & Jongman (1990) and Sereno (1994) found that high-frequency nouns exhibit a tendency to contain back vowels, while verbs show a bias toward front vowels. Since I am interested in lexicon-wide differences between nouns and verbs, I will test whether the reported difference in vowel distribution can also be found when tested on all noun and verb lemmas. I employ a scalar operationalization of the front/back difference, with [iː, ɪ, ɛ, æ,] assigned to the front category (coded 1) and [uː, ʊ, aː, ɐ, ɔː] to the back category (coded 5). Central vowels [ɜː, ʌ, ə] were coded (3). With diphthongs, the scores of the two vowels involved in their production were averaged, for example, the diphthong [a] was coded (3).

Variables 7-8 Average vowel backness and Average vowel height: Monaghan et al. (2005) find that not only the vowel quality of the tonic vowel, but also the average vowel quality of all vowels in a word differs between nouns and verbs in child-directed speech. Corresponding to the findings for the main stressed vowel, Monaghan et al. report a tendency for more back vowels in nouns as compared to verbs. Moreover, they find that nouns and verbs differ also with regard to vowel height, with nouns containing low vowels more often than verbs (see Durieux & Gillis 2001 for a similar result). Both of these tendencies will be tested here. In order to calculate the average height and backness of a words vowels, numerical scales for these dimensions were used. For vowel backness, the same scale as described for variable 6 was employed. Vowel height was coded as follows: High vowels [iː, ɪ, uː, ʊ] received a score of (1), low vowels [æ, aː, ɐ] a score of (5) and central vowels [ɜː, ɛ, ʌ, ə, ɔː] a score of (3). With diphthongs, the scores of the two vowels were averaged. For each word a mean score of all vowels was calculated, see the following examples for illustration.

Average vowel backness

(5) organ (back vowel and center vowel, (5+3)/2 = 4)
(6) fortitude (two back vowels and a front vowel, (5+5+1)/3 = 3.66)

Average vowel height
(7) technique (mid vowel and high vowel, (3+1)/2 = 2)
(8) pudding (two high vowels, (1+1)/2 = 1)

Variables 9-13 Distribution of consonant types : A number of differences regarding the distribution of certain consonant types have been observed between nouns and verbs. Kelly (1992) reports nouns to feature more nasal consonants than verbs (see also Durieux & Gillis 2001). Monaghan et al. (2005, 2007) find further differences in the distribution of coronal consonants, velars and approximants. In order to test for these differences, the proportion of the respective group of consonants in relation to all other consonants of the word were calculated. Words with no consonants were coded (0). Examples (9-13) exemplify the different consonant types and the calculation of the corresponding ratios.

(9) Ratio of nasals [n, m, ŋ], e.g. nose (0.5)
(10) Ratio of velars [k, g, ŋ], e.g. crate (0.33)
(11) Ratio of bilabials [b, p, m], e.g. mouth (0.5)
(12) Ratio of coronals [d,t,ʤ, ʧ, ð, θ, n, l, r, s, z], e.g. ditch
(13) Ratio of approximants [w,r,l,j], e.g. wade (0.5)

Variables 14-15 Initial bilabial and Initial approximant: Two differences have been uncovered that affect only the first segment of a word. In child-directed speech, Monaghan et al. (2007) find that verbs have a higher probability of beginning with an approximant, while nouns have a tendency to have a bilabial consonant as the first segment. Each word in the data was coded for whether it began with the relevant segment (1), or not (0).

3 Analysis of Data

3.1 Phonological differences between nouns and verbs

In a first step, I tested for statistically significant differences between the unambiguous noun and verb lemmas as pertains to the 15 aforementioned phonological variables. Since it is possible that the variables interact with morphological differences between nouns and verbs, I created a separate sample containing only monomorphemic nouns/verbs. Mann-Whitney U-tests were calculated separately for both that sample and the sample containing all words. The results of this analysis are displayed in Table 1.

Nr	Phonological Property	Nouns	Verbs	Z	r	p
1	Stress Pattern	0.32(0.58)	-0.34(0.13)	32.06(9.27)	0.21(0.13)	***(***)
2	Word Length	2.80(1.80)	2.5(1.38)	12.52(15.62)	0.14(0.22)	***(***)
3	Syllabic Complexity	3.46(3.41)	3.47(3.54)	0.69(12.56)	0.01(0.18)	n.s.(***)
4	Onset Complexity	1.01(1.11)	1.05(1.24)	-3.08(-6.23)	0.02(0.09)	n.s.(***)
5	Ratio reduced vowels	0.21(0.15)	0.16(0.13)	10.02(3.34)	0.06(0.05)	***(***)
6	Vowel backness - tonic syllable	2.38(2.67)	1.65(2.61)	19.06(0.46)	0.01	***(n.s.)
7	Vowel backness - word average	2.48(2.69)	2.39(2.61)	7.61(2.39)	0.05(0.03)	***(**)
8	Vowel height - word average	2.56(2.77)	2.42(2.73)	7.63(1.45)	0.05(0.02)	n.s.(***)
9	Ratio nasals	0.19(0.17)	0.16(0.13)	9.59(4.63)	0.11(0.07)	***(***)
10	Ratio velars	0.11(0.14)	0.09(0.13)	7.21(1.16)	0.07(0.02)	***(n.s.)
11	Ratio bilabials	0.17(0.19)	0.16(0.18)	3.85(1.77)	0.04(0.02)	***(+)
12	Ratio coronals	0.60(0.56)	0.63(0.57)	-8.98(1.39)	0.10(0.02)	***(n.s.)
13	Ratio approximants	0.19(0.20)	0.20(0.21)	-2.69(0.87)	0.03(0.01)	n.s.(***)
14	Initial approximant	0.12(0.10)	0.12(0.11)	0.02(-0.86)	0.01(0.01)	n.s.(n.s.)
15	Initial bilabial	0.23(0.25)	0.17(0.20)	7.69(3.49)	0.05(0.05)	***(***)

***=p<0.001 **=p<0.01 * p<0.05 +=p<0.1 n.s.=not significant

Table 1: Phonological differences between nouns and verbs in CELEX (results for monomorphemic words in parentheses)

Let me illustrate how the results in Table 1 can be interpreted by taking variable (1) Stress Pattern as an example. The Noun and Verb columns contain the average values of the variable Stress Pattern for the two word classes. The higher average values for nouns (both in the overall sample and in the monomorphemic sample) indicate that nouns have a greater tendency to be stressed on the initial syllable than verbs. The rightmost column, which contains the p-value, reveals that this difference is highly statistically significant. The column labeled r contains the results of a calculation of the point-biserial correlation coefficient, which serves as a measure of effect size. Values of 0.1/0.3/0.5 are typically interpreted as indicating small, moderate and large effects, respectively. For the variable Stress Pattern it can thus be concluded that we are dealing with a small to medium effect.

Overall, the results show that nouns and verbs exhibit statistically significant differences in many phonological dimensions. Of the fifteen variables tested, thirteen yield significant results for the sample of all nouns and verbs (ten for the sample of monomorphemic words). There are some conspicuous differences between the complete sample and the sample of monomorphemic words that indicate an effect of morphological complexity on phonological differences between nouns and verbs. While in both samples nouns emerge as longer than verbs, this effect is more pronounced in the monomorphemic sample (compare the correlation coefficient r for word length across the two samples in Table 1 above). Thus, the length effect is offset to a certain extent by morphology. A similar difference is found with regard to onset complexity. Word onsets of verbs are generally more complex than the ones of nouns, but this difference is greater with monomorphemic words. However, it is not generally true that phonological differences are attenuated in morphologically complex words. A number of distributional biases of different consonant types were found to be not significant in the sample of monomorphemic words, but in the overall sample (variables 10-13). The same observation can be made for two out of three variables associated with vowel quality (variables 6+8).

In order to obtain an initial idea of how these parameters apply to cases of conversion, I conducted the same analysis on the conversion sample. In doing so, I divided the sample into a Noun-to-Verb and a Verb-to-Noun part, based on the analysis of attestation dates from the *Oxford English Dictionary* (OED) carried out by Bram (2011).[8] Again, I conducted separate analyses for both the entire sample and for the monomorphemic portion. If phonological criteria are an indicator of original word class, then we can expect the same differences that exist between nouns and verbs in the CELEX data to also distinguish the VtoN from the NtoV part. The results of this analysis are displayed in Table 2.

Nr	Phonological Property	NtoV	VtoN	Z	*r*	p
1	Stress Pattern	0.71(0.82)	0.12(0.54)	9.15(9.15)	0.31(0.31)	***(**)
2	Word Length	1.68(1.45)	1.39(1.23)	8.12(5.88)	0.19(0.18)	***(***)
3	Syllabic Complexity	2.94(3.04)	3.23(3.36)	-7.59(-6.56)	0.18(0.20)	***(***)
4	Onset Complexity	1.15(1.18)	1.32(1.39)	-5.26(-5.63)	0.12(0.17)	***(***)
5	Ratio reduced vowels	0.15(0.12)	0.10(0.08)	5.21(2.27)	0.12(0.07)	***(***)
6	Vowel backness - tonic syllable	2.66(2.77)	3.03(3.19)	-5.69(-4.01)	0.13(0.12)	***(*)
7	Vowel backness - word average	2.45(2.69)	2.50(2.57)	-2.35(-1.54)	0.06(0.05)	*(n.s.)
8	Vowel height - word average	2.74(2.75)	2.65(2.70)	1.64(1.09)	0.04(0.03)	n.s.(n.s.)
9	Ratio nasals	0.19(0.17)	0.16(0.13)	9.59(4.63)	0.11(0.07)	***(**)
10	Ratio velars	0.12(0.13)	0.12(0.13)	0.73(0.26)	0.02(0.01)	n.s.(n.s.)
11	Ratio bilabials	0.19(0.20)	0.17(0.18)	1.82(2.20)	0.04(0.07)	+(n.s.)
12	Ratio coronals	0.55(0.55)	0.59(0.58)	-2.38(1.27)	0.06(0.04)	*(n.s.)
13	Ratio approximants	0.19(0.20)	0.27(0.28)	-5.43(4.25)	0.12(0.13)	***(***)
14	Initial approximant	0.12(0.12)	0.19(0.17)	-3.92(-2.55)	0.09(0.08)	***(*)
15	Initial bilabial	0.27(0.27)	0.21(0.21)	3.14(2.80)	0.07(0.08)	***(*)

***=p<0.001 **=p<0.01 * p<0.05 +=p<0.1 n.s.=not significant

Table 2: Phonological differences between NtoV and VtoN conversion cases (grouping based on historical criterion, results for monomorphemic words in parentheses)

In accordance with the CELEX analysis, the two groups of conversion cases also differ significantly along many phonological dimensions, with thirteen variables yielding significant results (ten in the sample of monomorphemic words). Comparing the results of the CELEX to the conversion sample reveals a large degree of overlap. Of the fifteen variables tested, thirteen exhibit the same tendencies across both groups. The only area of divergence is Vowel Backness, where unambiguous nouns in CELEX are characterized by a higher probability of back vowels than verbs, as concerns both the tonic syllable and also the overall average of all vowels. In contrast, in the conversion sample, VtoN cases exhibit a greater propensity for back vowels in the tonic syllable, and the results for the average of all vowels are mixed.

In summary, nouns and verbs differ along a large number of phonological dimensions, most of which also carry over to conversion cases. However, many of these differences are relatively subtle, as indicated by the predominantly rather low effect size values (see the results of the point-biserial correlation coefficient r in Tables 1-2, above).

3.2 Modeling phonological cues to word class using conditional inference trees and random forests

The large number of significant contrasts between nouns and verbs which are found to also distinguish NtoV from VtoN cases indicate that an analysis of phonological characteristics to determine directionality would provide promising results. However, in order to use one or more phonological variables in this way, it is necessary to investigate how reliable the criteria actually are in determining word class, in the sense that, if a word has the property X, it has a Y probability of being a noun or verb, respectively. Such information could then be used to determine the source word class in cases of conversion. What needs to be taken into consideration when conducting this investigation is that certain criteria apply only to subsamples in the data. This is the case, for example, with the variable Stress Pattern, as differences in stress position between nouns and verbs are restricted to words longer than one syllable.

What we would like to find out, of course, is, how reliable all cues and their combinations are when jointly tested. Testing the influence of multiple variables necessitates a multifactorial analysis of the aforementioned phonological characteristics. One methodological resource that is excellently suited for the task is that of Conditional Inference Trees, which rely on a partitioning mechanism that creates subsamples of the data. This is extremely useful for the current purpose, as it means that the method can take into account that some phonological properties are relevant for certain parts of the lexicon only. Conditional inference trees employ a recursive splitting algorithm that uses the predictor variables to divide the data into binary subsamples in an iterative fashion. This means that the algorithm first splits the dataset into two subsamples by using the phonological variable that most strongly distinguishes the data with regard to classifications of the response variable (= word class). It will then test the resultant two subsamples as to whether other variables are significantly associated with the response variable and if they are, impose further binary splits. The algorithm continues until no further statistically significant associations between the predictor variables and the response variable are found in any of the subsamples (see Tagliamonte & Baayen 2011, Wiechmann & Kerz 2011 for descriptions and applications of this method; for more general information on classification trees, see Breiman et al. 1993).

Before applying this technique, careful consideration was given to which sample should be used. What needed to be taken into account was that the share of words that undergo conversion has certain properties that distinguish it from the lexicon overall. Therefore, using the entire CELEX noun-verb lemma sample would have been problematic. One issue was that the CELEX sample is heavily skewed toward nouns (88% nouns, 12% verbs), while in the conversion sample the ratio of cases derived from nouns (NtoV) is only 67% (based on OED data). By using the unfiltered CELEX data, one would therefore run the risk of overestimating the frequency of nouns, and by extension NtoV cases. A further difference between the conversion sample and CELEX overall is the considerably higher ratio of monomorphemic words in the former. This presented a problem, since the analysis above revealed that some of the phonological parameters differ between monomorphemic and multimorphemic words. In order to take into account these aspects, I created a random stratified sample from the CELEX data that matches the sample of conversion data by Bram (2011) with regard to noun-verb ratio and morphological complexity (see Table 3).[9]

	Morphologically simplex	Morphologically complex	Total
Nouns	2,119	696	2,815
Verbs	1,017	334	1,351

Table 3: CELEX lemma sample modeled after the sample of conversion cases

I then calculated a conditional inference tree based on this sample, using the CTREE function of the Partykit package in R (see Hothorn Zeileis 2015). The result of that calculation is a fairly complex tree, see Figure 1.

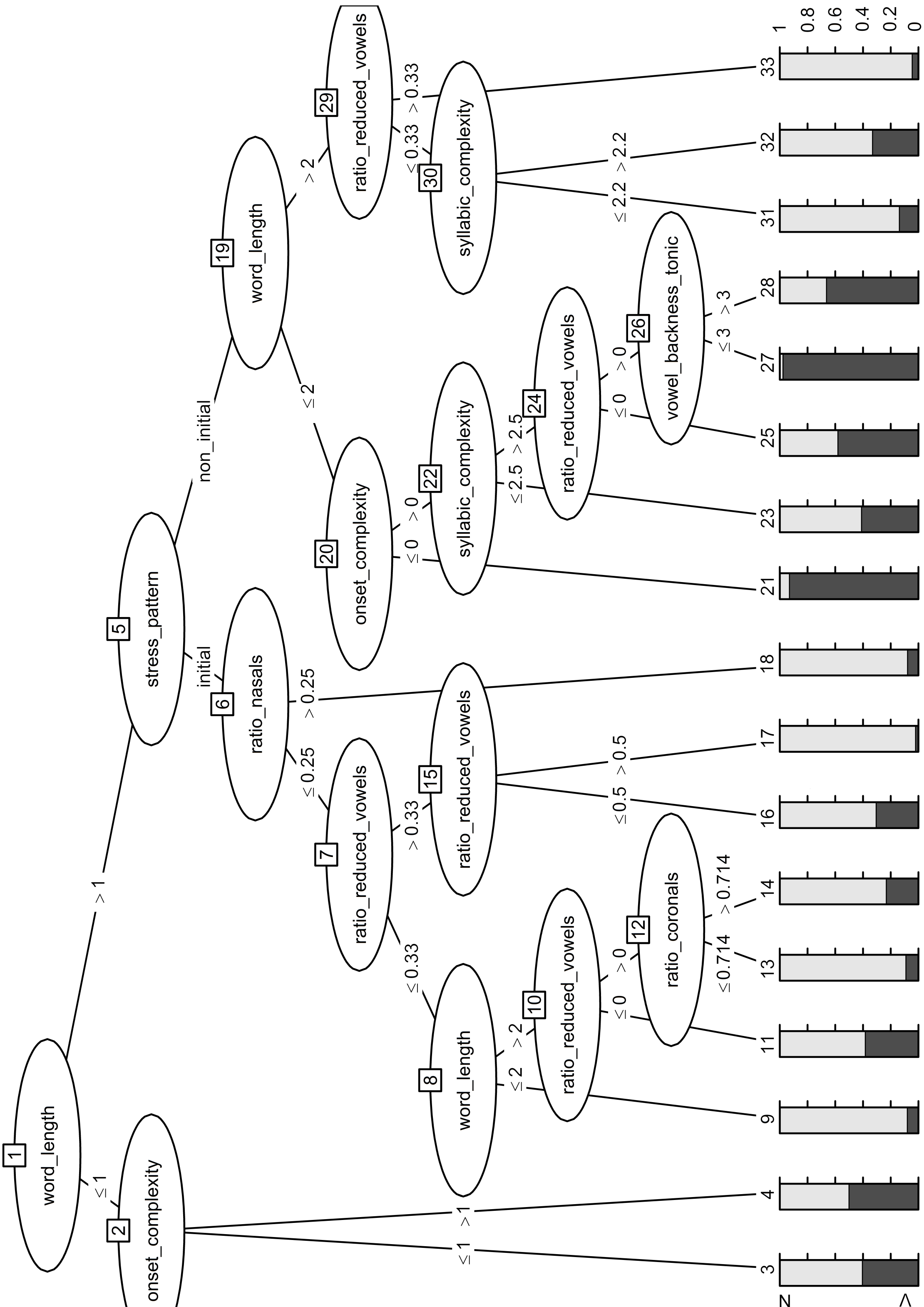

Figure 1: Conditional Inference Tree (the bicolored areas at the bottom of the tree signify the noun-verb ratios of the respective subsamples, light gray = noun ratio, dark gray = verb ratio)

When inspecting the tree, the first thing to notice is that nine out of the fifteen variables tested feature in the tree, thus are found to significantly distinguish between nouns and verbs. These are Word Length, Stress Pattern, Word Onset Complexity, Ratio of reduced vowels, Syllabic Complexity, Vowel backness of the tonic syllable, Ratio of nasals and Ratio of coronals. The variables featuring in the tree impose sixteen splits in the data (see labeled nodes in Figure 1), resulting in a total of seventeen terminal nodes at the bottom of the tree (also called 'leaves'). These terminal nodes represent subsamples that are not divided any further; these subsamples have varying noun-verb ratios, as represented through the ratio of light vs. dark gray areas in the columns at the bottom of Figure 1. Overall, the tree correctly predicts word class for 72.1% of all cases (C=0.63). This is a rather modest result, given that a null-model would arrive at an accuracy of 67% by merely guessing noun for every word (as this is the ratio of nouns in the sample). However, when inspecting the leaves of the tree in Figure 1, we can see that the tree makes much better predictions for some subsamples than for others. This can be gleaned from the very different ratios of light gray (noun ratio) vs. dark gray (verb ratio) areas in the terminal nodes of the tree, which indicate that the predictive accuracy is not 72.1% for all words but varies considerably between different subsets in the data (see the exact noun-verb ratios of all terminal nodes in Table 4 in the Appendix).

Let us take a closer look at some details of the tree in order to obtain a better idea of its specific predictions. In doing so it makes sense to follow the different branches of the tree from top to bottom. The first split of the tree (node 1) is brought about by the variable Word Length, dividing the sample into the two subsamples of monosyllables and longer words. Following the left branch, the monosyllables, shows that there is a further split of that branch (node 2), which indicates that monosyllabic words with at least one consonant in the onset have a higher probability of being verbs (node 4) than those with an unfilled onset (node 3). However, the verb ratio in both of these terminal nodes is close to 50% (59% and 50%, respectively). In other words, while there is a statistically significant difference in onset complexity, the phonological characteristics do not allow for a very accurate distinction between verbs and nouns in the sample of monosyllables.

A second, larger subsample in the data is split off by node 5, which separates the polysyllabic share of the data into a subsample containing words with main stress on the first syllable and a subsample with main stress elsewhere. Starting with the former, it can be gleaned from Figure 1 that a number of variables further subdivide this part of the tree (see nodes 6, 7, 8, 10, 12 and 15, below node 5). However, all of the terminal nodes of that branch are biased toward nouns (with at least 60% noun ratio), so that we can conclude that disyllabic and longer words with main stress on the initial syllable are predominantly nouns.

Returning to the top levels of the tree, the next important split (node 19) divides the non-initially stressed words into disyllabic and longer words. Words longer than two syllables are indicated by the rightmost branch of the tree, which exhibits further splits imposed by the variables Ratio of reduced vowels and Syllabic Complexity. However, all terminal nodes of that branch lean toward nouns (nodes 31, 32, 33). The left branch starting at node 19 encompasses the disyllabic words with main stress on the final syllable. For current purposes, this part of the tree is the most interesting one, as that branch features terminal nodes with varying noun-verb ratios. Four out of five terminal nodes are biased toward verbs (nodes 21, 25, 27, 28), while one node (node 23) exhibits a noun-bias.

Concluding, while the overall tree structure is very complex, it is possible to identify subsamples that are characterized by clear biases toward one or the other word class. I will return to this finding in section 3.3, when discussing how the results can be used to determine directionality in cases of conversion.

As has been shown in the literature on conditional inference trees, results based on an individual tree can be problematic, as it may reflect configurations of the predictor variables in the given sample that are not representative of the population as a whole (see Strobl et al. 2009), a problem known as overfitting in statistics (see e.g. Baayen 2008). This means that the tree may include splits that reflect a random characteristic of the sample that would disappear by introducing only small changes to the data. One way to resolve this situation is to calculate a so-called RANDOM FOREST, which is an aggregation of many CONDITIONAL INFERENCE TREES, and then check which patterns in the data are borne out by a majority of the trees. Such an analysis allows for greater generalizability and can also help to test the validity of the initial classification tree, as it reveals whether the variables featuring in that tree also yield significant results in an aggregation of many trees. I therefore calculated a random forest of 500 trees using the CFOREST function of the PARTY package in R (Hothorn et al. 2015). Each tree of the forest was built on a sample randomly generated from the original sample.[10] One way of using the results of this calculation is to let the random forest predict word class based on the phonological variables. This is achieved by doing a vote of the individual trees for each data point and results in a predictive accuracy of 76% (C=0.79), which represents an improvement over the single conditional inference tree. A further important insight can be obtained from the random forest by calculating each variables contribution to its overall accuracy. The bar plot (Figure 2) visualizes the result of such a calculation (which was carried out by employing the VARIMP function of the PARTY package in *R*).

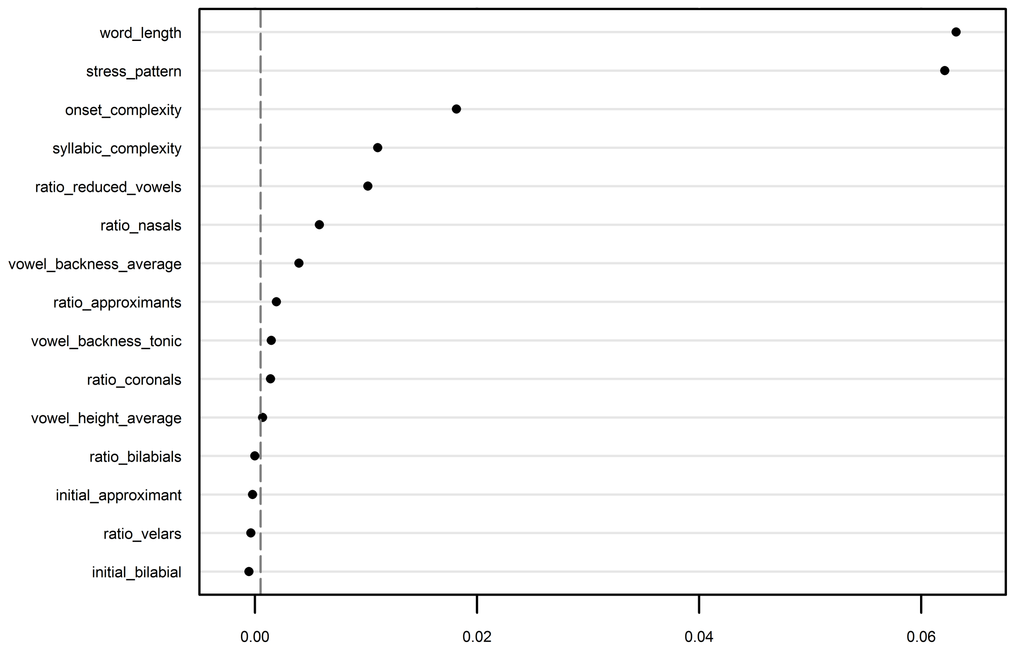

Figure 2: Variable importance of predictor variables in a random forest of 500 conditional inference trees

Variables in Figure 2 are ranked by importance, with the variables at the top representing the most important predictors of word class. The dashed line serves as an indicator of the relevance of the variables, with those variables whose dots are positioned to its right representing statistically significant predictors (cf. Strobl et al. 2009, Shih 2011).[11] When comparing the results of this analysis to the initial tree, a number of correspondences can be detected. First, all variables that feature in the tree are also identified as relevant predictors by the random forest, which can be interpreted as a validation of the initial tree as it indicates that the tree is most likely not overfitting the data. Second, the order of the variables importance in Figure 2 is reflected in the order of splits in the tree, in the sense that the first splits are imposed by those variables that are identified as the most important variables in the random forest (cf. Figure 1, where Word Length and Stress Pattern impose the first splits of the tree).

In summary, the calculation of the conditional inference tree and the random forest show that a number of phonological variables contribute significantly to predicting word class in the given sample. However, the accuracy of predictions varies considerably across different subsamples of the data. This point will be taken up in the following.

3.3 Using phonological cues to determine directionality in conversion

There are several ways in which the results of the empirical analysis can facilitate making informed decisions about directionality in cases of conversion. The most obvious one is to let the tree or forest make binary predictions for the conversion sample in order to identify which cases it puts in the noun and which in the verb category. The advantage of doing this is that one would obtain predictions for all possible conversion cases, as the tree/forest makes predictions for all words, and thus for all conversion cases. However, the conditional inference tree reveals that the phonological variables are involved in complex interactions, with many variables distinguishing between nouns and verbs only in selected subsamples, resulting in drastically different predictive accuracies for these subsamples. Using the entire tree means to lose that information and to settle for an average rate of 72% accuracy, thereby also using predictions that are much less reliable since the accuracy is far below that value for some shares in the data (see Table 4 in the Appendix, which contains the word class ratios of all terminal nodes of the conditional inference tree).

Another option is to restrict oneself to those parts of the lexicon for which phonological properties are considerably accurate predictors of word class, as evidenced by high purity ratios of the corresponding terminal nodes in the tree. The results for these subsamples should therefore allow for more reliable predictions with regard to directionality. For example, instead of using the entire tree, one could focus on those terminal nodes whose predictive accuracy exceeds 80%. This would mean employing only the predictions of the terminal nodes 9, 11, 17, 18, 27, 31, and 33 (see also Table 4 in the Appendix). The increase in accuracy of this solution comes with the drawback that these nodes cover only 23.5% of the conversion sample by Bram (2011).

Both alternatives, either using the entire tree or solely selected terminal nodes, involves taking into account the fairly complex interactions that are indicated by the individual nodes of the tree. Thus, applying these phonological cues practically requires either access to the statistical model or at least to this paper; their application is thus complex and time-consuming. Therefore, I propose a third, simpler alternative here, which makes use of the general patterns in the data revealed by the conditional inference tree and translates these into simple-to-apply rules of thumb. In doing this I identify those parts of the tree that share certain phonological characteristics and are characterized by a clear bias toward one of the two word classes. While this operation results in some information loss, the advantage is that easily applicable guidelines can be formulated that still ensure reasonable accuracy.

The first result of this procedure is that directionality cannot be reliably determined in monosyllabic words on the basis of phonological criteria (56% nouns, 44% verbs). As mentioned above, monosyllabic nouns and verbs differ significantly with regard to their onset complexity (see node 2 in Figure 1). However, the predicted probability for either word class arrived at by considering this difference is between 50% and 59%, an accuracy that is not much better than mere guessing. This is a disappointing finding, as 52% of the words in the sample of conversion cases by Bram (2011) are monosyllabic words. Thus for a substantial share of the data, phonological characteristics are in fact of little help. However, for the remainder of conversion cases, three robust cues can be identified, which are listed in the following.

(i) Disyllabic and longer words that have main stress on the initial syllable are derived from nouns (82.8% nouns)

This rule corresponds to the well-known tendency of polysyllabic nouns to be stressed on the initial syllable and is arrived at by lumping together the terminal nodes 9, 10, 12, 15, 16, 17, and 18 of the conditional inference tree.

(ii) Trisyllabic and longer words are derived from nouns (80.6% nouns)

This cue is obtained by lumping together the right-hand branch of the tree, which contains words that are not stressed on the first syllable and are longer than two syllables (nodes 30, 32, and 33) and adding those words longer than two syllables that are stressed on the initial syllable. With 80.6% noun ratio in this subsample, this is a fairly accurate cue to determine directionality, and it is also one that can be applied straightforwardly, as all that is needed is a syllable count.

(iii) Disyllabic words with stress on the final syllable are derived from verbs, except for those with a filled word onset and a syllabic complexity of less than 2.5 phonemes per syllable (83.1% verbs)

This rule carves out the verb share of the sample and is extracted from the left branch starting at node 19, adjoining the terminal nodes 21, 26, 27, and 28, but excluding node 23.

In summary, I have outlined three alternative ways of applying the results of the conditional inference tree to conversion data. These three alternatives are briefly repeated here along with their scope of application.

a) Phonological cues based on the entire conditional inference tree: source word class is determined by using the binary predictions of the conditional inference tree when applied to the conversion dataset (applies to 100% of the conversion data).

b) Simple phonological cues: source word class is determined by applying the rules of thumb laid out in (i-iii) (applies to 45.5% of the conversion data).

c) Fine-grained phonological cues: source word class is determined by relying on the cues derived from the terminal nodes of the conditional inference tree that exceed 80% accuracy (applies to 23.5% of the conversion data).

Choosing between these three alternatives means a tradeoff between accuracy and scope: the most accurate (and at the same time most complex) set of cues (c), applies to only a limited share of the data, while using the entire tree allows for predictions in 100% of the cases (see a), however, these predictions are often of poor accuracy. The simplified phonological cues (b), as specified in (i)-(iii), seem to strike a good balance between accuracy and scope, as they can be applied to 45.5% of the data (almost all words except monosyllables) and still reach reasonably high accuracy values. In the following, the results of these alternative ways of applying phonological criteria will be tested on the sample of conversion cases and compared against other criteria to determine directionality.

3.4 Phonological cues in the context of other criteria

In applying these cues to the conversion dataset, there is no way in which their true correctness can be calculated, as there is no other criterion that could determine source word class with complete certainty. However, as an indirect test of the accuracy of predictions, I will calculate the agreement rate of the phonological cues with other criteria commonly used to determine directionality. For that calculation, I use the entire dataset of noun-verb conversion cases by Bram (N=1,880) and rely on his operationalization and coding (see Bram 2011: 116-157 for details). The criteria used for comparative purposes are listed in the following.

I) Historical criterion: the word whose use is attested first is the source word (attestation dates taken from the Oxford English Dictionary (2011)).

II) Frequency: the word more frequently instantiated in corpus data from the British National Corpus (online version as provided by Davis 2004) is the source word.

III) Semantic Dependence: if the definition of the core sense of one word includes the other word in the Merriam-Websters Third Unabridged Dictionary (2007), the first is considered to be the derived word.

IV) Semantic Range: the word with more word senses as listed in the Merriam-Websters Third Unabridged Dictionary (2007) is the source word.

Before comparing the phonological cues to these criteria, it is necessary to discuss whether different measures of directionality can and should be compared to each other at all. It is obvious from the list above that the individual criteria capture very different aspects of directionality. Given these differences, an important question is whether these criteria measure what Bram (2011) terms a coherent notion or idea of directionality. I understand this question to mean whether the individual criteria are compatible with each other and whether they could be regarded as different measurements of the same multi-faceted notion of directionality. This is both a theoretical and an empirical question. Regarding the theoretical aspect, one may ask whether any of the aforementioned criteria exclude each other on logical grounds, such that a certain result obtained for one criterion leads to a conflicting result or the inapplicability of another criterion. This does not seem to be the case, as it is certainly conceivable that a particular instance of conversion is characterized by all aforementioned criteria yielding the same result. A case in point is the example of *bottle* (N) / *to bottle* (V), as all criteria listed above yield the result of NtoV conversion: according to the analysis by Bram (2011), the noun bottle predates the verb *to bottle* by 266 years, the noun *bottle* is considerably more frequent in corpus data, the noun *bottle* has a larger number of senses (as listed in dictionaries), and there is also a semantic dependence in the direction ofNtoV (as a bottle is needed to bottle something). Interestingly, also the phonology would yield this result, as an initially stressed disyllabic word indicates the source word should be a noun (cf. section 3.3). The correspondence between the individual criteria can be understood to be the result of a conversion process in which, diachronically, a new categorically specified word is derived (via different semantic patterns as described in Marchand 1963 or Clark & Clark 1979), which may show a dependence on the semantics of the source word. During this process usually only a limited number of the senses of the source word are mapped to the derived word (Plank 2010), which may then explain its restricted semantic range and lower frequency of use. The derivation of the verb *to bottle* from the noun *bottle*, therefore, seems to instantiate a coherent process that is marked by a correspondence between the individual criteria. However, skepticism has been voiced as to whether this is typically the case. For example, Sanders (1988: 171) doubts that there is a systematic general correspondence between the historical criterion and the synchronic derivational relation in conversion pairs. This skepticism is fueled by discussions of individual examples in which semantic changes [...] overwrite the original direction of conversion (Plag 2003: 108). It is an empirical question whether such cases are typical of conversion word pairs or merely exceptions to an overall pattern of correspondence between the criteria.

This question is investigated by Bram (2011), who tested a large number of cases as to whether the different criteria exhibit a tendency to conspire, thus yield the same result, or whether they are typically in conflict with one another. For the criteria listed in in (I-IV), he finds a mean agreement rate of 67.9%, which he interprets in favor of correspondence and therefore as evidence of a coherent concept of directionality. This result shows, however, that there is still a fair amount of disagreement or uncertainty among the individual criteria. According to Bram (2011: 321), this is not a problem of the criteria themselves, but a result of the less-than-ideal data sources that are available for operationalizing them, a point I mention above (see section 1). In any event, the agreement ratio significantly exceeds chance agreement, so that even using less-than-ideal data sources, evidence is obtained of systematic correspondences between the criteria. Thus, along with Bram (2011) I take this result as indicating that the criteria measure different aspects of the same coherent concept of directionality that typically marks the word-formation process of conversion. This means that these criteria can be used as a standard of comparison for the phonological cues: if we assume the phonological criteria to capture an additional aspect of this same process, we would also expect them to exhibit a tendency to agree with the other criteria. It is this hypothesis that is tested in the following.

The bar plot in Figure 3 illustrates the agreement rates of the three ways to use phonological cues to determine directionality as laid out above (a-c in section 3.3), when compared to the criteria in I-IV (see above in this section).[12] It also contains information about the mean agreement rate of the criteria used for comparative purposes (see I-IV), which is represented by the dashed line (the solid line marks 50% chance agreement).

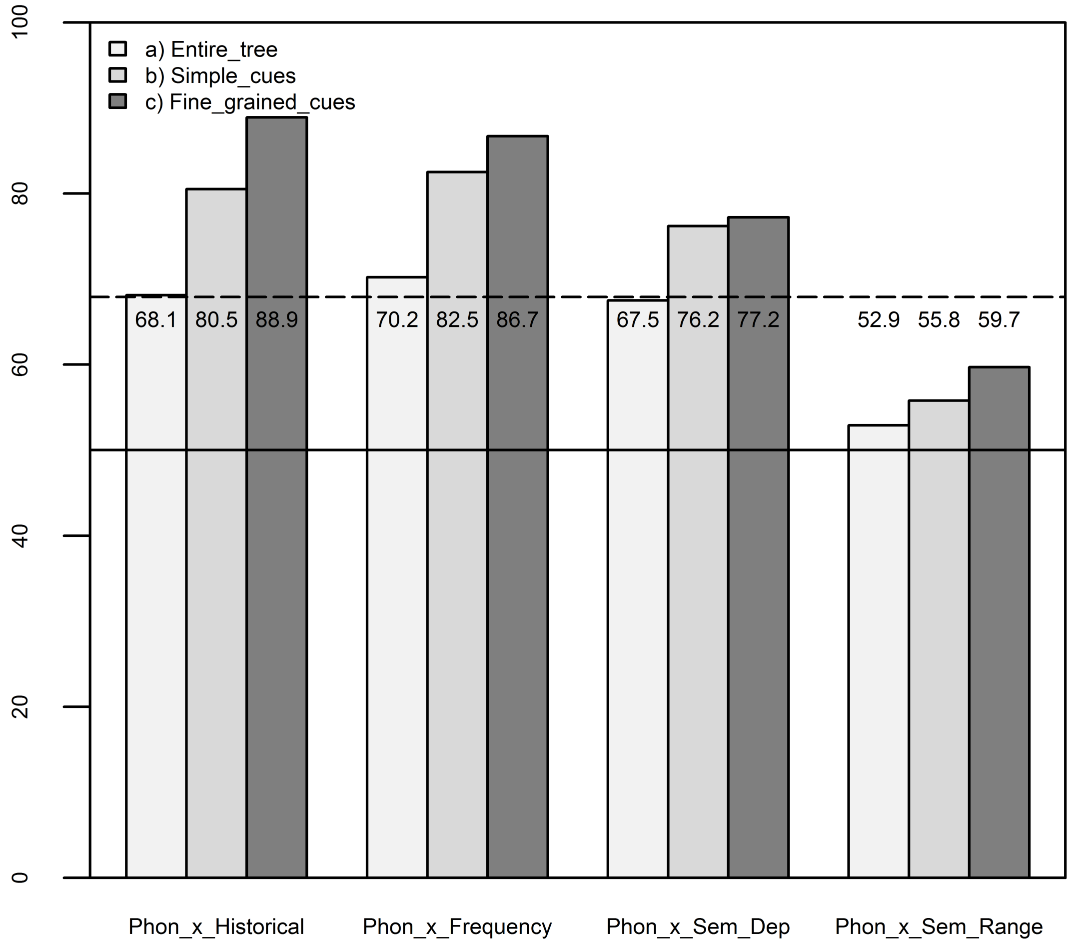

Figure 3: Agreement rates of phonological cues with other criteria commonly employed to determine directionalit

The most important result of this calculation is that, overall, agreement of the phonological cues with the other criteria is fairly high (see height of the individual bars in Figure 3). When averaging over all pairwise comparisons, the agreement rate between phonological cues and the other criteria is 72.2%, which is higher than the average agreement rate among the four other criteria. Albeit in an indirect way, this result lends support to the general idea of using phonological cues to determine directionality, as it indicates that the phonological cues measure an additional aspect of the same coherent concept of directionality discussed above.

When comparing the three different ways to use phonological cues (see a-c in section 3.3, marked by the color of the bars), a clear pattern emerges. Irrespective of which criterion is used for comparison, there is a rise in agreement rate from the predictions of the entire tree (a), to the simple cues (b), to the fine-grained cues (c). This is an important result, as it means that differences in the accuracy of the cues identified through the conditional inference tree correspond to differences in agreement with the other criteria. This correspondence again points to a coherence between the phonological cues and the other criteria and, furthermore, corroborates the approach undertaken, namely to rely on accuracies obtained via a conditional inference tree in order to determine directionality.

A further interesting result that can be gleaned from the bar plot is that agreement rates vary according to the criterion the phonological cues are compared to. Agreement rates are high with the historical[13] and the frequency criterion, while slightly lower with Semantic Dependence, and barely exceeding the 50% chance agreement rate with Semantic Range. This result will be discussed further below.

4 Discussion

This paper has sought to empirically assess to what extent phonological properties of word class can be employed to determine directionality in noun-verb conversion. This aim was approached in two steps: First, I investigated the distribution of a large number of phonological properties between nouns and verbs in the entire lexicon as represented in CELEX. Second, these phonological properties were tested as to how well they work as cues to word class and therefore as a criterion for the linguist to determine directionality in cases of conversion.

The first step, the analysis of the CELEX lemma sample, reveals that the distribution of many phonological properties differs significantly between nouns and verbs (see Tables 1 and 2). This is an important result as it shows that many phonological characteristics of word classes, previously evidenced only in child-directed speech (cf. Monaghan et al. 2005, 2007), are in fact properties of the lexicon as a whole. This finding expands results obtained by Durieux Gillis (2001), who provide evidence of some phonological differences between nouns and verbs in CELEX data but tested a considerably smaller set of variables.

Through the calculation of a conditional inference tree and an ensemble of such trees (a *random forest*), a number of variables and their combinations that may serve as cues to word class were identified. I discussed three ways to employ these results to inform decisions about directionality. These three alternatives differ with regard to their scope and accuracy (see a-c in section 3.3). The first one, using the predictions of the entire tree (a), is characterized by a rather low accuracy and correspondingly results in low agreement rates with other criteria used to determine directionality (see Figure 3). This is mostly due to the fact that the tree also makes predictions for the sample of monosyllabic words, for which it was found that word class cannot be reliably determined using phonological properties. This finding represents a major achievement of the multifactorial method applied, as it shows that lumping together monosyllables with longer words masks important differences in cue strength of phonological properties between the two groups. These differences can be explained by the fact that the potential of phonological variability between nouns and verbs is smaller in monosyllables. The variables that most strongly discriminate between nouns and verbs, namely Word Length and Stress Pattern, do not vary in monosyllabic words and therefore cannot distinguish the two word classes in this share of the data. The same is true of the relatively important variable, Ratio of reduced syllables, as the syllable in monosyllabic content words cannot be reduced. Other variables, targeting the complexity of the syllable or the occurrence of certain segment types could theoretically differ between monosyllabic nouns and verbs, but are found to not distinguish word class in this part of the sample. The only exception is *Word Onset Complexity*, whose cue strength is, however, negligible.

In contrast to employing predictions of the entire tree, the approaches (b) and (c) apply to smaller shares of the lexicon, yet their predictions are considerably more accurate. This is mostly due to their being restricted to words of at least two syllables in length, a subsample for which phonological cues tend to be fairly accurate. When applied to cases of conversion, both (b) and (c) result in fairly high agreement rates with the other criteria (see Figure 3). These two ways to use phonological cues therefore seem to be appropriate additions to the linguists toolkit for determining directionality. While (c) requires complex analyses, since it relies on the predictions of the individual terminal nodes of the conditional inference tree, (b) is a collection of rules that can be applied straightforwardly.

Both (b) and (c) represent significant improvements over the more intuitive applications of phonological properties to determine directionality that are discussed in the literature; a general improvement is that the results obtained allow for the application of phonological properties whose cue strength has been precisely quantified. Furthermore, previously undetected phonological variables and their interactions that may serve as cues could be identified. An important result in that regard is that the variable Word Length can be used as a reliable criterion in informing decisions about directionality, as, irrespective of other phonological characteristics, words longer than two syllables have a high probability of being nouns (see (iii) in section 3.3). The difference in length between nouns and verbs was first discovered by Cassidy & Kelly (1991), but it had not been mentioned in the literature on conversion yet. The results obtained furthermore confirm the validity of the phonological variable of stress position for disyllabic noun-verb pairs, which has been suggested as a criterion to determine directionality in previous research (see Marchand 1969, Kiparsky 1982, 1997 and Balteiro 2007). The conditional inference tree reveals important interactions of this variable with others, showing that disyllabic words with final stress that have a filled word onset but a low overall syllabic complexity are in fact more likely to be nouns than verbs. This indicates that not all disyllabic words with final stress can be reliably categorized as VtoN conversion.

The general question that remains to be addressed is what is truly gained by adding phonological cues as a further criterion to determine directionality. From a qualitative perspective, having a new criterion that is based on a different logic than the already available ones is a gain in itself, especially since the phonological criterion suffers less from operationalization problems than the other criteria (see discussion in section 1). A further benefit would be an improvement that was also measurable quantitatively. A true gain in this sense would be the determination of directionality in cases for which other criteria are inapplicable. Phonological cues can contribute to the resolution of such cases. For example, of the share of 81 conversion cases in the Bram (2011) dataset for which the criterion Semantic Dependence is not informative, 53 can be determined using the simple phonological rules (b), as described in section 3.3. There is also a fair share of conversion pairs for which the historical development cannot be reconstructed with certainty (see discussion in Bram 2011: 211-231). Given the close correspondence between phonological cues and the historical criterion, the former can be taken as an indicator of the original word class and thereby provide evidence of the probable diachronic development.

The close correspondence between these two criteria raises the general question of why the agreement rates vary depending on which of the other criteria the phonological criterion is compared to. The rather high agreement between phonology and the diachronic development can be explained by word-class specific phonological characteristics of the source word that are retained in the derived word in the diachronic conversion process. There is considerably less correspondence between the phonological criterion and the semantic criteria. One explanation for this finding may be the aforementioned changes of the semantics of words, which may lead to an incongruence between the historical criterion and the semantic criteria (see Plag 2003: 108 for a discussion of examples). Since the form of words is not affected by changes in semantics and therefore still reflects the historical derivational relation, the same incongruence can then be observed with the phonological criterion. A further reason for a comparably low correspondence between phonology and semantics may lie in the difficult operationalization of the semantic criteria. The agreement rates calculated here rely on Brams (2011) operationalization, which are based on dictionary definitions. As mentioned above, these definitions are characterized by a high frequency of nouns, which may lead to an inflation of NtoV classifications (Bram 2011: 297, 314).

The correspondence between the phonological criteria and the frequency criterion is comparatively high, close to the agreement rates for the historical criterion. Differences in frequency are usually explained by a wider range of meanings of the source form, which is therefore more frequently used (cf. section 1). This explanation would suggest that also the correspondence between phonology and the criterion of semantic range should be high, which, however, is not the case. In fact, even the agreement rate between frequency and semantic range is rather low (57.7%). These results once again suggest that the operationalization of the semantic criteria via dictionary definitions may not accurately capture the usage of a word in naturally occurring language. Other than that no convincing explanation can be offered for why frequency and phonology show a high correspondence, while semantic range does not.

A further question regarding the place of phonological properties compared to the other criteria is that of its possible cognitive underpinnings. While the perspective of this paper is that of the linguist who seeks to determine the source word, it is an interesting question whether the criteria used to determine directionality capture aspects of how the two words are related to each other cognitively. This is most obviously the case with the semantic dependence criterion, as this criterion asks whether the meaning of one word requires the knowledge of the concept denoted by the other word of the pair (for example, understanding the meaning of the verb *to bottle* requires the concept denoted by the noun *bottle*). Also the criteria frequency of use and semantic range are related to the mental representation of the respective words, as a large number of senses and a high frequency of use strengthen the representation of a word in the mental lexicon. The question is whether with the phonological criteria a similar connection to the mental representation of the two words can be drawn. In previous research, evidence has been provided of speakers having (implicit) knowledge of phonological characteristics of nouns and verbs. It has been shown that such characteristics affect speakers when categorizing words as either nouns or verbs (Sereno & Jongman 1990), influence whether speakers use novel forms as either nouns or verbs (Cassidy & Kelly 1991, 2001), and impact speakers when making up novel nouns or verbs (Hollmann 2012, 2013). These results demonstrate that phonological characteristics are registered in the mental lexicon with the two word classes, possibly in the form of schemas as suggested by Hollmann (2013). Thus, it can be argued that the kind of directionality measured by the phonological criteria has a representational correlate, as language users may associate the form of a word more strongly with the noun or the verb representation by comparing it to stored schemas of these word classes. In this sense phonological cues nicely complement the other criteria that are also related to the mental representation of conversion cases. Thinking of the word pairs as two different but related signs, the semantic criteria capture the meaning part of the signs, the frequency criterion captures their overall representational strength, and the phonological cues capitalize on the phonological form of the signs and their relation to prototypical word-class schemas.

5 Conclusion

This paper has empirically demonstrated that directionality in noun-verb conversion can be co-determined via phonological properties of word class. Phonological cues can be employed for words of at least two syllables in length, either as an additional criterion or as a substitute for other, more commonly applied criteria to determine directionality.

The findings for noun-verb conversion raise the question of whether phonological cues can also be applied to conversion between other word classes. A logical candidate to consider is adjectives, as conversion between adjectives and nouns, and between adjectives and verbs is well attested. However, it has been shown that, while more similar to nouns than verbs, adjectives are positioned between these two word classes as pertains to a number of important phonological dimensions (Berg 2000). For this reason, it is likely that the overall phonological differences between adjectives and noun/verbs, respectively, are considerably smaller than the ones between nouns and verbs, which would reduce their validity as cues to word class. Hence, phonological properties seem to be a less promising resource to use on cases of conversion involving adjectives. Conversion may also happen between function and content words, an example being the nominal *ifs and buts*, and phonological differences between function and content words are well documented (e.g., Monaghan et al. 2005). However, since conversion involving function words seems to be a minor process (Bram 2011: 67-70) for which the direction of conversion is fairly clear, employing the criterion of phonological cues does not seem to be necessary.

Beyond the word-formation process of conversion, the present study has implications for the study of phonological properties of word classes more generally, as many phonological differences between nouns and verbs previously attested only in child-directed speech are found to be a characteristic of the lexicon as a whole. Furthermore, the present analysis has brought to light complex interactions between phonological properties through the method of conditional inference trees, which show that many differences between nouns and verbs exist only in certain parts of the lexicon. Such fine-grained information is immensely important for the understanding of the phonological characteristics of word classes. In obtaining it, the multifactorial analysis carried out here represents a major improvement over the monofactorial testing of phonological differences in previous research (e.g., Durieux & Gillis 2001, Monaghan et al. 2005).

Notes

[1]This paper is dedicated to Hans Marchand and his impressive contribution to the study of conversion. Marchand, like the author of the present lines, taught at Bard College and must have enjoyed the same beautiful views of the Hudson River. I wish to thank Ingo Plag for sharing his thoughts on the role of phonology in conversion with me and Thomas Berg for providing valuable comments on an earlier version of this paper. Furthermore, I wish to express my gratitude to the FunCog Team and the audiences at the 41st *Österreichische Linguistiktagung* and the 48th Annual Meeting of the *Societas Linguistica Europae* for helpful feedback. All remaining errors are mine. I gratefully acknowledge funding for this study by the *Deutsche Forschungsgemeinschaft* (grant LO 2135/1-1).

[2]The abbreviations N and V are used throughout this paper to indicate the word classes noun and verb, respectively.

[3]Kiparsky (1997: 15) discusses this point in detail. He argues that there is a difference between a true dependence and what may be regarded as only a seeming semantic dependence. Of particular interest to this argument are cases such as *hammer*, which may be categorized as denominal verbs, but which do not exhibit a true dependence according to Kiparsky, because it is possible to hammer something without a hammer, by for example using a shoe to do the hammering (cf. Kiparsky 1997: 17). Contra Kiparsky (1997) it could be argued that a hammer can still be regarded as the prototypical instrument that is used for hammering and it would be the instrument that one would assume to be used if there was no information to indicate otherwise. Therefore, one may still assume a semantic dependence in the direction of noun to verb. See Sanders (1988: 174) for a discussion of this argument.

[4]It is unclear how disyllabic noun-verb pairs with a consistent iambic stress pattern, e.g. comMAND are handled by Kiparskys (1982) theory of lexical phonology, since due to their verbal origin, his model would predict these to be stress-shifting. Through stating that nouns [...] formed from [...] verbs **may** shift to the nominal stress pattern [emphasis is mine], Kiparsky (1982: 12) leaves room for these exceptions, but his model does not offer an explanation for why these cases escape the stress shift (see also Bram 2011: 42 on this point). Incidentally, similar problems hold for the view of regarding stress shift as 'internal modification' that marks the derivational relationship between the two words, as under this view one would expect all disyllabic noun-verb pairs to be stress-shifting (see Bauer 2008: 200-201).

[5]A possible additional phonological variable is discussed by Marchand (1964: 16). He identifies a number of typical word endings which, he argues, indicates derivation from noun to verb and verb to noun, respectively, and which he subsumes under the label phonetic shape of the word. Despite the label, however, the endings suggested, *ment*, *ure* and *(a)tion*, capture morphological information of words, as these endings are well-known noun-deriving suffixes (cf. Bauer, Lieber & Plag 2013: 196-202). For this reason, this variable is not considered in the current analysis.

[6]A similar analysis based on CELEX data has been carried out by Durieux & Gillis (2001), however, their study was not informed by recent results from research on child-directed speech,

and they tested a smaller number of phonological parameters.

[7]I used the online-accessible version WebCELEX at www.celex.mpi.nl (Baayen et al. 2001).

[8]Bram (2011) uses the online version of the OED to extract attestation dates. Since he does not specify when exactly he accessed the OED, I refer to it as the Oxford English Dictionary (2011) in this paper, as Bram completed his thesis in 2011.

[9]The morphological status of each word in CELEX is classified as follows: M= monomorphemic, C= complex word, U = undetermined, O = obscure, R = may include a root, Z= zero-derivation. One problem in creating a sample of unambiguous nouns and verbs from CELEX that matches the Bram (2011) sample with regard to morphological status is of course the group of Z-cases, which make up 17.2% of the Bram (2011) sample. Except for their being classified as instances of conversion by the CELEX editors, the label Z is not informative with regard to the morpholocial complexity of these words. Therefore, I manually coded all words in the Z-group, deciding for each word whether it was monomorphemic (M), or morphologically complex (C) (the large majority of them are monomorphemic words). After this additional coding step, I used the breakup of the different classes in the Bram (2011) sample for creating the CELEX sample (see Table 3).

[10]The parameter *mtry* of the CForest function, which determines the number of input variables randomly sampled as candidates at each node, was set to 15, that is, equal to the total number of predictor variables. This setting enables bootstrap aggregating.

[11]The dotplot was created using R code provided by Shih (2011). The dashed line marks the absolute value of the lowest negative-scoring variable, identified through the calculation of variable importance. It can be taken as an indication of whether a given variable is empirically relevant or not, as [t]he rationale for this rule of thumb is that the importance of irrelevant variables varies randomly around zero (Strobl et al. 2009: 342).

[12]In calculating agreement rates, I took into account only the data for which both criteria that are compared make predictions, excluding those cases for which one or both cannot be applied.

[13]This result corresponds to findings by Balteiro (2007: 166-7), who reports a high agreement between using stress position in disyllabic words and the historical criterion to determine directionality in N-V conversion.

References

Baayen, Rolf H.; Piepenbrock, R. and L. Gulikers. 2001. WebCelex. Online resource; http://celex.mpi.nl.

Baayen, Rolf H. 2008. Analyzing linguistic data: A practical introduction to statistics using R. Cambridge: Cambridge University Press.

Balteiro, Isabel 2007. The directionality of conversion in English: A dia-synchronic study (LINGUISTIC INSIGHTS 59). Frankfurt am Main, Berlin, Bern, New York, Paris, Wien: Lang.

Barner, David & Alan Bale 2002. No Nouns, No Verbs: Psycholinguistic Arguments in Favor of Lexical Underspecification. Lingua: International Review of General Linguistics 112(10): 771-791.

Bauer, Laurie. 2008. Derivational Morphology. Language and Linguistics Compass 2: 196-210.

Bauer, Laurie, Rochelle Lieber & Ingo Plag 2013. The Oxford Reference Guide to English Morphology. Oxford: Oxford University Press

Berg, Thomas 2000. The position of adjectives on the noun-verb continuum. English Language and Linguistics 4(2): 269-293.

Bram, Barli 2011. Major total conversion in English: The question of directionality. PhD thesis. University of Wellington.

Breiman, Leo, Jerome Friedman, Charles Stone & R.A. Olshen 1993. Classification and regression trees (THE WADSWORTH STATISTICS, PROBABILITY SERIES). New York: Chapman Hall.

Cassidy, Kimberly W. & Michael H. Kelly 1991. Phonological information for grammatical category assignments. Journal of Memory and Language 30: 348-369.

Cassidy, Kimberly W. & Michael H. Kelly 2001. Children's use of phonology to infer grammatical class in vocabulary learning. Psychonomic Bulletin Review Journal 8(3): 519-523.

Clark, Eve V. & Herbert H. Clark 1979. When Nouns Surface as Verbs. Language 55(4): 767-811.

Davies, Mark. 2004. BYU-BNC. (Based on the British National Corpus from Oxford University Press). Available online at http://corpus.byu.edu/bnc/.

Don, Jan 2004. Categories in the Lexicon. Linguistics: An Interdisciplinary Journal of the Language Sciences 42(5): 931-956.

Don, Jan 2005. Roots, deverbal nouns and denominal verbs. In Geert Booij, Emiliano Guevara, Angela Ralli, Salvatore Sgroi & Sergio Scalise (eds.), Morphology and linguistic typology: Proceedings of the fourth Mediterranean Morphology Meeting (MMM4). Catania: Univ. di Catania Facolt di Lettere e Filosofia. 91-103.

Don, Jan & Marian Erkelens 2008. Possible phonological cues in categorical acquisition: evidence from adult categorization, Studies in Language 32(3): 670-682.

Durieux, Gert & Stevens Gillis 2001. Predicting grammatical classes from phonological cues: An empirical test. In Jrgen Weissenborn & Barbara Hhle (eds.), Approaches to bootstrapping: Phonological, lexical, syntactic, and neurophysiological aspects of early language acquisition. Amsterdam/Philadelphia: John Benjamins. 189-230.

Farrell, Patrick 2001. Functional shift as category underspecification. English Language and Linguistics 5: 109-130

Halle, Morris & Alec Marantz 1993. Distributed morphology and the pieces of inflection. In Kenneth L. Hale & Samuel J. Keyser (eds.), The view from building 20: Essays in linguistics in honor of Sylvain Bromberger (CURRENT STUDIES IN LINGUISTICS 24), 111-176. Cambridge, Mass: MIT Press.

Hollmann, Willem B. 2012. Word classes: Towards a more comprehensive usage-based account. Studies in Language 36(3): 671-698.

Hollmann, Willem B. 2013. Nouns and verbs in Cognitive Grammar: Where is the sound evidence? Cognitive Linguistics 24(2): 275-308.

Hothorn, Torsten & Achim Zeileis. 2015. partykit: A Toolkit for Recursive Partytioning, R package (I employed version 0.8-0).

Hothorn, Torsten, Kurt Hornik, Carolin Strobl & Achim Zeileis. 2015. party: A Laboratory for Recursive Partytioning, R package (I employed version 1.0-15).

Jespersen, Otto 1946. A modern English grammar on historical principles: Pt. 6 Morphology. Copenhagen: Munksgaard.

Kelly, Michael H. 1992. Using sound to solve syntactic problems. Psychological Review 99(2): 349-364.

Kelly, Michael H. 1996. The role of phonology in grammatical category assignments. In James L. Morgan & Katherine Demuth (eds.), Signal to syntax: Bootstrapping from speech to grammar in early acquisition. Mahwah (NJ): Lawrence Erlbaum Associates Publishers. 249-262.

Kelly, Michael H. & J. Kathryn Bock 1988. Stress in time. Journal of Experimental Psychology: Human Perception and Performance 14(3): 389-403.

Kiparsky, Paul. 1982. From cyclic phonology to lexical phonology. In Harry van der Hulst & Norval Smith (eds.), The structure of phonological representations: Part 1. Dordrecht/Cinnaminson: Foris Publications. 131-176.

Kiparsky, Paul. 1997. Remarks on denominal verbs. In Alex Alsina (ed.), Complex predicates (CSLI LECTURE NOTES 64). Stanford (CA): CSLI Publ. 473-500.

Marantz, Alec. 1997 No Escape From Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon. In Alexis Dimitriadis et al. (eds.), Proceedings of the 1998 Penn Linguistics Colloquium (UNIVERSITY OF PENNSYLVANIA WORKING PAPERS IN LINGUISTICS: VOL. 4.2). 201-225.

Marchand, Hans 1963. On a Question of Contrary Analysis, with Derivationally Connected but Morphologically Uncharacterized Words. English Studies: A Journal of English Language and Literature 44: 176-187.

Marchand, Hans 1964. A set of criteria for the establishing of derivational relationship between words unmarked by derivational morphemes. Indogermanische Forschungen 69(1): 10-19.

Marchand, Hans. 1969. The Categories and Types of Present-day English Word-formation: A Synchronic-Diachronic Approach. 2nd ed. Munich: Beck.

Merriam-Webster Unabridged Dictionary. 2007. Online version. http://unabridged.merriam-webster.com

Monaghan, Padraic, Nick Chater & Morten H. Christiansen 2005. The differential role of phonological cues in grammatical categorisation. Cognition 96: 143-182.

Monaghan, Padraic, Morten H. Christiansen Nick Chater 2007. The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. Cognitive Psychology 55: 259-307.

Monaghan Padraic & Morten H. Christiansen 2008. Integration of multiple probabilistic cues in syntax acquisition. In Heike Behrens (ed.), Trends in corpus research: Finding structure in data (TILAR SERIES) Amsterdam: John Benjamins. 139-163.

Nazzi, T. D. M. Houston. 2006. Finding verb forms within the continuous speech stream. In Kathy Hirsh-Pasek (ed.), Action meets word: How children learn verbs. Oxford: Oxford Univ. Press. 64-87.

Oxford English Dictionary. Online version. 2011. Oxford: Oxford University Press.

Plag, Ingo. 1999. Morphological productivity: structural constraints in English derivation (TOPICS IN ENGLISH LINGUISTICS 28). Berlin/New York: de-Gruyter.

Plag, Ingo. 2003. Word-formation in English (CAMBRIDGE TEXTBOOKS IN LINGUISTICS). Cambridge: Cambridge University Press.

Plank, Frans. 2010. Variable Direction in Zero-Derivation and the Unity of Polysemous Lexical Items. Word Structure 3(1): 82-97.

R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sanders, George. 1988. Zero-derivation and the Overt Analogue Criterion. In Michael T. Hammond & Michael Noonan (eds.), Theoretical morphology: Approaches in modern linguistics. San Diego: Academic Press. 155-175.

Sandoval, Michelle; Gonzales, Kalim & Rebecca Gmez. 2012. The road to word class acquisition is paved with distributional and sound cues. In Patrick Rebuschat & John N. Williams (eds.), Statistical learning and language acquisition (STUDIES IN SECOND AND FOREIGN LANGUAGE EDUCATION). Berlin/New York: deGruyter. 145-170.

Sereno, Joan A. 1994. Phonosyntactics. In L. Hinton, J. Nichols & J. Ohala (eds.), Sound symbolism. Cambridge: Cambridge University Press. 263-275.

Sereno, Joan A. & Allard Jongman 1990. Phonological and Form Class Relations in the Lexicon. Journal of Psycholinguistic Research 19(6): 387-404.

Shih, Stephanie. 2011. Random forests for classification trees and categorical dependent variables: an informal Quick Start R Guide. www.stanford.edu/ stephsus/R-randomforest-guide.pdf.

Štekauer, Pavol. 1996. A theory of conversion in English. Frankfurt am Main, Berlin, Bern, New York, Paris, Wien: Lang.

Strobl, Carolin, James Malley & Gerhard Tutz 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods 14(4): 323-348.

Tagliamonte, Sali A. & Rolf H. Baayen 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. Language Variation and Change 34: 135-178.

Wiechmann, Daniel & Elmar Kerz 2013. The positioning of concessive a verbal clauses in English. Assessing the importance of discourse-pragmatic and processing-based constraints. English Language and Linguistics 17(1): 1-23.

Zipf, George K. 1945. The meaning-frequency relationship of words. Journal of General Psychology 33: 251-256.

A Appendix

Terminal Node	Noun ratio	Verb ratio
3	0.593	0.407
4	0.496	0.504
9	0.916	0.084
11	0.613	0.387
13	0.906	0.094
14	0.765	0.235
16	0.691	0.309
17	0.976	0.024
18	0.918	0.082
21	0.066	0.934
23	0.586	0.414
25	0.416	0.584
27	0.020	0.980
28	0.333	0.667
31	0.859	0.141
32	0.665	0.335
33	0.951	0.049

Table 4: Noun and verb ratios of the terminal nodes of the conditional inference tree (see Figure 1)