# Predicting the Direction of Derivation in English Conversion

**Max Kisselew**[*], **Laura Rimell**[†], **Alexis Palmer**[‡], and **Sebastian Padó**[*]
[*] IMS, Stuttgart University, Germany
{pado,kisselmx}@ims.uni-stuttgart.de
[†] Computer Laboratory, University of Cambridge, UK
laura.rimell@cl.cam.ac.uk
[‡] Leibniz ScienceCampus, ICL, Heidelberg University, Germany
palmer@cl.uni-heidelberg.de

## Abstract

*Conversion* is a word formation operation that changes the grammatical category of a word in the absence of overt morphology. Conversion is extremely productive in English (e.g., *tunnel*, *talk*). This paper investigates whether distributional information can be used to predict the diachronic direction of conversion for homophonous noun–verb pairs. We aim to predict, for example, that *tunnel* was used as a noun prior to its use as a verb. We test two hypotheses: (1) that derived forms are less frequent than their bases, and (2) that derived forms are more semantically specific than their bases, as approximated by information theoretic measures. We find that hypothesis (1) holds for N-to-V conversion, while hypothesis (2) holds for V-to-N conversion. We achieve the best overall account of the historical data by taking both frequency and semantic specificity into account. These results provide a new perspective on linguistic theories regarding the semantic specificity of derivational morphemes, and on the morphosyntactic status of conversion.

## 1 The Morphology of Conversion

Word formation operations that change the grammatical category of a word in the absence of overt morphology pose interesting linguistic challenges. Such operations are highly productive in English, especially between the categories of noun and verb (consider examples such as *tunnel* or *walk*, also more recent examples such as *email*). This phenomenon is also observed in other languages, for example German (Vogel, 1996), Dutch (Don, 1993), Hungarian (Kiefer, 2005), and Bulgarian (Manova and Dressler, 2005). We call these cases "conversion", without any theoretical commitment.

Conversion in English, especially N-to-V conversion, has been extensively studied within morphology and syntax. Historically, accounts for this phenomenon involved (a) *conversion* (proper), a category-changing word-formation operation assumed to be different from other types of derivational morphology (Koziol, 1937; Bauer, 1983; Plag, 1999), or (b) *zero-derivation*, involving a derivational affix akin to *-ize, -ify, -er*, etc., but which happens to be phonologically null (Bloomfield, 1933; Marchand, 1969; Kiparsky, 1982). Most current syntactic theories of conversion, though, are based on *underspecification*. This is the idea that meaning-sound correspondences are stored in the lexicon as uncategorized roots and obtain a grammatical category when combined with syntactic heads, which may be phonologically null or overt (Halle and Marantz, 1993; Pesetsky, 1995; Hale and Keyser, 2002; Borer, 2005a; Arad, 2005).

The range of meanings resulting from conversion has been a key source of evidence for various theoretical approaches. Verbs derived via N-to-V conversion can have a wide variety of meanings, seemingly constrained only by the template 'action having to do with the noun', such as in the phrase *celluloid the door open*, meaning 'use a credit card to spring the lock open' (Clark and Clark, 1979). V-to-N conversion has a narrower semantic range. It is likely to result in a noun referring to the event described by the verb or to its result, e.g. *talk* (Grimshaw, 1990).

This paper presents a computational study of conversion: we study which factors are able to account for *diachronic* precedence in cases of English V-to-N and N-to-V conversion. The goal is to predict, e.g., that *tunnel* was originally used as a noun and *walk* as a verb. Historical precedence provides a theory-neutral ground truth which we treat as a proxy for the actual direction of conversion.



*Proceedings of the 14th Annual SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 93–98,
Berlin, Germany, August 11, 2016. 

We use methods from distributional semantics to test two morphological hypotheses: (1) that derived forms are less frequent than their bases, and (2) that derived forms are more semantically specific than their bases. We use information theoretic measures to gauge semantic specificity, applying these measures for the first time to theoretical questions regarding derivational morphology.

## 2 The Direction of Derivation

We analyze corpus data as a source of evidence for the direction of derivation in lemmas attested in both nominal and verbal contexts. We take historical precedence as the gold standard for the grammatical category of the base. For example, the lemma *tunnel* was first attested in English as a noun around the year 1440, and as a verb in 1577, according to the Oxford English Dictionary. The gold standard direction of derivation is therefore N-to-V. On the other hand, the lemma *drive* was first attested as a verb around 900 and as a noun in 1697, so the gold standard direction is V-to-N.

The idea of predicting a direction of derivation is not uncontroversial from a theoretical perspective. According to underspecification accounts of conversion, there is no direction to predict, since both nominal and verbal uses result from syntactic categorization of a root which is unspecified for category. Nevertheless, even underspecification allows for the fact that some roots seem to be used primarily in one category or another (Harley and Noyer, 1999; Arad, 2005; Borer, 2005b). Moreover, historical precedence provides an objective ground truth, regardless of any particular theory of word formation (Rimell, 2012).

**Gold Standard.** Our gold standard consists of 1,044 English lemmas which have undergone N-to-V conversion and 948 lemmas which have undergone V-to-N conversion. We obtained the historical precedence data from CELEX (Baayen et al., 1995) using the WebCelex interface.[1] N-to-V lemmas are coded in CELEX as monomorphemic nouns and also conversion verbs; V-to-N lemmas are monomorphemic verbs and also conversion nouns.[2] We limited the dataset to lemmas which are monomorphemic in their base grammatical category in order to avoid root compounds and denominal verbs formed from already-derived nouns, such as *leverage* and *commission*, which we believed would complicate the analysis. We manually excluded a handful of lemmas which appeared in both CELEX searches due to polysemy.

[1] http://celex.mpi.nl

[2] N-to-V lemmas have a CELEX entry with Class=N (part of speech) and MorphStatus=M (monomorphemic), and a second entry with Class=V and MorphStatus=Z (conversion). The converse holds for V-to-N lemmas.

## 3 Methods

### 3.1 Hypotheses

We advance two main hypotheses which can be investigated in a corpus-based fashion.

1. Derived forms are *less frequent* than their base words.

2. Derived forms are *semantically more specific* than their base words.

The first hypothesis is not entirely straightforward, since some derived forms are in fact more frequent than their bases, especially when the base is frequent (Hay, 2001). However, derived forms have been found to have lower frequency than their bases in general (Harwood and Wright, 1956; Hay, 2001); therefore, while the hypothesis as stated may be an oversimplification, we use it as a first approximation to a frequency-related analysis of conversion.

The second hypothesis corresponds to the Monotonicity Hypothesis of Koontz-Garboden (2007), which states that derivational morphemes always add content to their bases, in the form of compositional lexical semantic operators. Linguistic support for this proposal comes from cross-linguistic examination of word-formation operations, such as causative and anticausative operations on verbs. If conversion is the result of a phonologically null derivational affix, we would expect the Monotonicity Hypothesis to hold. Semantic specificity, or complexity – with a derived word assumed to have a more complex, or narrower, meaning because it has more semantic subcomponents – has also been used as a diagnostic for the direction of derivation in conversion (Plag, 2003), but based on linguistic judgments rather than distributional semantics.

The rest of this section is concerned with operationalizing these two hypotheses. We first describe the corpus that we are using, then the semantic representations that we construct from it to pursue the second hypothesis, and finally the concrete predictors that instantiate the hypotheses.

### 3.2 Corpus

Our corpus is a concatenation of the lemmatized and part-of-speech (PoS) tagged BNC[3] and ukWaC corpora[4], containing 2.36 billion tokens. Both corpora are widely used for building distributional spaces. Together, they cover a large range of text types both in terms of genres and of domains.

Since we will use this corpus to extract information about the noun and verb usages of morphologically unmarked conversion cases, it is a pertinent question how well standard part-of-speech taggers recognize this distinction. To test this, we carried out a manual annotation study.

From each corpus we extracted 33 examples each of 100 lemmas, chosen randomly from the lemmas in the gold standard, half on the N-to-V conversion list and the other half on the V-to-N list. Two of the authors, both native English speakers, annotated the PoS tags for correctness. Inter-annotator agreement was $\kappa$=0.68. Overall, the accuracy of the PoS tags was 85%, which we considered sufficiently reliable for good quality category-specific representations.[5]

While many lemmas and their instances were straightforward to annotate as either noun or verb, some examples presented difficulties. Two prominent cases were gerunds and adjectival forms (***forked*** *tongue*, *fuselage was* ***skinned*** *with aluminum*), although there were a variety of other, less frequent cases. In these instances we used the overt inflectional morphology as a guide; for example, *-ing* or *-ed* endings indicated a verb. This strategy is based on the fact that inflectional morphology strictly selects for the part of speech of its base.

### 3.3 Vector Representations

To measure semantic specificity, we perform a distributional analysis which represents each conversion case with two 10,000-dimensional bag-of-words vectors: one for the verb and one for the noun, relying on automatic PoS tags (cf. Section 3.2). The dimensions correspond to the most frequent content words in the corpus. The context window size is set to 5 words on either side of the target. Following standard practice, we apply a Positive Pointwise Mutual Information (PPMI) transformation and L1-normalize each vector.[6]

**Downsampling.** The use of vectors based on co-occurrence counts poses a methodological difficulty, because word frequency is a potential confounder for the information-theoretic measures with which we operationalize the specificity hypothesis (Section 3.4). The potential difficulty arises because more frequent words might have denser vectors (more non-zero values), which could lead to observing spurious increases in specificity that are merely correlates of frequency rather than the result of a conceptual shift. To avoid this danger, we balance the frequencies of bases and derived forms by *downsampling*. For each verb-noun conversion pair, both vectors are constructed from the same number of occurrences, namely $min(f_N, f_V)$, by skipping instances of the more frequent category uniformly at random. For example, *tunnel (n.)* occurs 38,967 times in the corpus and *tunnel (v.)* 2,949 times. Through downsampling, the vectors both for *tunnel (n.)* and for *tunnel (v.)* are constructed from 2,949 instances.

### 3.4 Operationalizing the Hypotheses

**Frequency.** We assess the frequency hypothesis by directly comparing the number of nominal and verbal corpus occurrences of a target lemma.

**Semantic Specificity.** We operationalize the semantic specificity hypothesis by applying measures of information content to distributional representations. This follows the example of two recent studies. In the context of hyponymy identification, Santus et al. (2014) proposed *entropy* as a measure of the semantic specificity $S(w)$ of a word $w$, via its distributional, L1-normalized vector $\vec{w}$. Entropy is supposed to be inversely correlated with semantic specificity, since higher specificity corresponds to more restrictions on context, which means lower entropy, defined as

$$S(w) = H(\vec{w}) = -\sum_i \vec{w}_i \cdot \log \vec{w}_i \qquad (1)$$

[3] `http://www.natcorp.ox.ac.uk`, tagged with the CLAWS4 tagger and the C5 tagset

[4] `http://wacky.sslmit.unibo.it`, tagged with TreeTagger and the Penn Treebank tagset

[5] We performed the same annotation on Wikipedia data, tagged with TreeTagger and the Penn Treebank tagset, but found the automatic PoS tagging to be less reliable. Therefore, we excluded it from consideration.

[6] Much recent work in distributional semantics has made use of low-dimensional, dense vectors, obtained either by dimensionality reduction of co-occurrence vectors, or as word embeddings from a neural network trained to optimize context prediction. Although reduced vectors and embeddings perform well on a variety of Natural Language Processing tasks, they are not suitable for our approach, because their feature weights are not interpretable probabilistically, which information-theoretic measures rely on.

| Predictor | N-to-V | V-to-N | all |
|---|---|---|---|
| Most Freq. Class | 100% | 0% | 52.4% |
| Entropy $H$ | 50.1% | 75.5% | 62.2% |
| KL divergence | 53.8% | **76.7%** | 64.6% |
| Frequency | **84.7%** | 58.7% | 72.3% |
| Freq + $H$ + KL | 77.4% | 76.0% | **76.8%** |

Table 1: Accuracies for predicting the direction of derivation, presented by gold standard direction (all results on downsampled space)

| Predictor | Estimate | Std. Err. | Sig. |
|---|---|---|---|
| Intercept | 0.15 | 0.06 | ** |
| Δ entropy | -2.08 | 0.18 | *** |
| Δ KL divergence | -2.22 | 0.18 | *** |
| Δ log frequency | 1.74 | 0.09 | *** |

Table 2: Logistic regression model (Δ always denotes noun value minus verb value)

The second study (Herbelot and Ganesalingam, 2013) was directly interested in measuring specificity and proposed to equate it with the Kullback-Leibler (KL) divergence $D$ between a word vector $\vec{w}$ and the "neutral" vector $\vec{n}$:

$$S(w) = D(\vec{w}||\vec{n}) = \sum_i \vec{w}_i \cdot \log \frac{\vec{w}_i}{\vec{n}_i} \qquad (2)$$

where $\vec{n}$ is the prior distribution over all words. We compute $\vec{n}$ as the centroid of approximately 28,000 word vectors in our vector space; the vectors are computed according to the procedure in Section 3.3. In this approach, higher KL divergence corresponds to higher semantic specificity. We note that the entropy and KL divergence values are closely related mathematically and highly correlated in practice ($\rho = 0.91$).

**Combined Model.** Finally, we combine the individual indicators (standardized differences in log frequency, entropy, and KL divergence within each pair) as features in a logistic regression model. We also experimented with including an interaction between the information-theoretic terms and frequency, but did not obtain a better model fit.

## 4 Results and Discussion

**Assessing the Hypotheses.** The quantitative results of our experiments are shown in Table 1. Compared against the most frequent class baseline, which assigns the most frequent direction in the gold standard — that is, N-to-V— to all cases, both our hypotheses are substantially, and significantly, more successful in predicting the direction of derivation (at a significance level of $\alpha = 0.05$).

Furthermore, the frequency hypothesis is more successful than the semantic specificity hypothesis on the complete conversion dataset. However, there is a striking complementarity between the frequency and specificity hypotheses with respect to the gold standard direction (N-to-V vs. V-to-N). Among the N-to-V cases, frequency excels with almost 85% accuracy, while the specificity predictors are at baseline level. V-to-N shows the opposite behavior, with above 75% accuracy for the specificity predictors and a sharp drop for frequency.

This complementarity among the predictors also enables the regression model to combine their respective strengths. It yields accuracies of 76%+ for both N-to-V and V-to-N conversion, with an overall accuracy of 76.8%, significantly better than frequency only. Table 2 shows normalized coefficients obtained for the four predictors in the model, all of which contribute highly significantly. Positive coefficients increase the log odds for the class N-to-V. As expected, the frequency difference between noun and verb comes with a positive coefficient: a higher noun frequency indicates N-to-V. According to hypothesis (2), we would expect a negative coefficient for the KL divergence difference between noun and verb and and a positive one for entropy. While this expectation is met for KL divergence, we also see a negative coefficient for entropy. This is due to the very strong correlation between the two predictors: the regression model uses the weaker one, entropy, as a "correction factor" for the stronger one, KL divergence.

Table 3 shows some examples (taken from the top of the alphabetically-ordered list of conversion cases), cross-classified by whether two complementary predictors (frequency and KL divergence) both make the correct prediction, disagree, or both fail, together with the number of conversion instances for each class. The two predictors agree more often than they disagree, but among the disagreements, the strong asymmetry between N-to-V (top) and V-to-N (below) is readily visible.

**Part-of-speech Differences as a Confounder.** A possible criticism of our results is that they arise primarily from distributional differences between

| correct direction: N-to-V | | | |
|---|---|---|---|
| wrong in both size=112 | wrong in $f$ size=48 | wrong in KL size=370 | correct in both size=514 |
| *augur* | *balk* | *age* | *air* |
| *biff* | *calk* | *alarm* | *alloy* |
| **correct direction: V-to-N** | | | |
| wrong in both size=132 | wrong in $f$ size=259 | wrong in KL size=91 | correct in both size=466 |
| *ally* | *account* | *act* | *accord* |
| *answer* | *address* | *babble* | *ache* |

Table 3: Conversion examples cross-classified according to the frequency ($f$) and KL divergence (KL) predictors, with sizes of various classes.

the two parts of speech (nouns and verbs) and are not specifically related to conversion. To test this hypothesis, we first inspected the means of entropy, KL divergence and log frequency in our sample and found that downsampling was successful in largely eliminating differences at the part-of-speech level (e.g., $\bar{H}_N$ = 6.57, $\bar{H}_V$ = 6.62). We tested the importance of the remaining differences by re-running our experiments with predictors that were normalized by part-of-speech (i.e., either subtracting the part-of-speech mean or dividing by it). The performance of the individual predictors hardly changed, nor did the performance of the logistic regression model (slight increase in accuracy from 76.8% to 77.0%). Our conclusion is that the patterns that we observe are indeed reflections of semantic shifts due to conversion, rather than inherent differences between parts of speech.

**Consequences of Asymmetry for Theory.** The asymmetry observed between N-to-V and V-to-N conversion in Table 1 suggests that different theoretical accounts of conversion may be appropriate for the two directions. The failure of the specificity hypothesis to predict the direction of N-to-V conversion at better than chance level is consistent with an underspecification approach, rather than a derivational one (cf. Section 1). The theoretical justification for our hypothesis (2), namely that derived forms are more semantically specific than their bases, assumes that the input to N-to-V conversion is a noun, to which semantic content is added in the form of a phonologically null operator. If, instead, an uncategorized root merges with a categorizing head to form both the noun and the verb, there is no reason why one would be more semantically specific than the other. On the other hand, the high accuracy of the information theoretic measures on V-to-N conversion are consistent with a derivational approach.

There is an interesting positive correlation between semantic regularity and (gain in) frequency. As often noted in the literature, the semantics of N-to-V conversion is irregular, with conversion verbs exhibiting a wide range of meanings – for example, *age*, meaning something like ‘increase in age’. In N-to-V conversion, the derived word often occurs less frequently than its base, possibly because the high level of semantic flexibility encourages nonce formations. On the other hand, V-to-N conversion has much more regular semantics, where the noun typically names the event or its result – for example, an *address* involves the act of *addressing*. In V-to-N conversion, frequency is a poor predictor of the direction of derivation, indicating that the derived word often occurs more frequently than its base, possibly because semantic regularity allows usages to become entrenched.

## 5 Conclusion

In this paper, we have analyzed the phenomenon of diachronic direction of derivation in English conversion. An initial experiment has shown a striking complementarity in the ability of frequency and semantic specificity to account for the direction of conversion in N-to-V and V-to-N cases, as well as good overall accuracy for a combined model. This opens up interesting avenues for future exploration. We believe corpus-based, distributional measures can yield useful insights for theoretical approaches to morphology and syntax. Finally, we note that Herbelot and Ganesalingam (2013) found a frequency-based measure and KL divergence to perform about equally well on the task of predicting lexical specificity, e.g. that *cat* is more specific than *animal*. The relationship between various corpus-based measures remains to be fully explored.

**Acknowledgments.** We thank the reviewers for their valuable comments and student assistant Olga Chumakova for her support. MK and SP acknowledge partial funding by Deutsche Forschungsgemeinschaft (SFB 732, Project B9). LR acknowledges EPSRC grant EP/I037512/1 and ERC Starting Grant DisCoTex (306920). AP acknowledges Leibniz Association grant SAS-2015-IDS-LWC and the Ministry of Science, Research, and Art of Baden-Württemberg.

## References


Maya Arad. 2005. *Roots and patterns: Hebrew morpho-syntax*. Springer, Dordrecht.

Harald R. Baayen, Richard Piepenbrock, and Leon Gulikers. 1995. *The CELEX lexical database. Release 2. LDC96L14*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Laurie Bauer. 1983. *English word-formation*. Cambridge University Press, Cambridge.

Leonard Bloomfield. 1933. *Language*. Henry Holt, New York.

Hagit Borer. 2005a. *In name only: Structuring sense, Volume I.* Oxford University Press, Oxford.

Hagit Borer. 2005b. *The normal course of events: Structuring sense, Volume II.* Oxford University Press, Oxford.

Eve V. Clark and Herbert H. Clark. 1979. When nouns surface as verbs. *Language*, 55:767–811.

Jan Don. 1993. *Morphological conversion*. Ph.D. thesis, Utrecht University.

Jane Grimshaw. 1990. *Argument Structure*. MIT Press, Cambridge.

Ken Hale and Samuel Jay Keyser. 2002. *Prolegomenon to a theory of argument structure*. MIT Press, Cambridge.

Morris Halle and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In S.J. Keyser and K. Hale, editors, *The view from Building 20*, pages 111–176. MIT Press, Cambridge.

Heidi Harley and Rolf Noyer. 1999. State-of-the-article: Distributed Morphology. *GLOT*, 4:3–9.

Frank W. Harwood and Alison M. Wright. 1956. Statistical study of English word formation. *Language*, 32(2):260–273.

Jennifer Hay. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics*, 39:1041–70.

Aurélie Herbelot and Mohan Ganesalingam. 2013. Measuring semantic content in distributional vectors. In *Proceedings of ACL*, pages 440–445, Sofia, Bulgaria.

Ferenc Kiefer. 2005. Types of conversion in hungarian. In *Approaches to Conversion/Zero-Derivation*. Waxmann, Münster.

Paul Kiparsky. 1982. Word formation and the lexicon. In Fred Ingeman, editor, *Proceedings of the Mid-America Linguistics Conference*, page 3–29. University of Kansas.

Andrew Koontz-Garboden. 2007. *States, changes of state, and the Monotonicity Hypothesis*. Ph.D. thesis, Stanford University.

Herbert Koziol. 1937. *Handbuch der englischen Wortbildungslehre*. C. Winter, Heidelberg.

Stela Manova and Wolfgang U. Dressler. 2005. The morphological technique of conversion in the inflecting-fusional type. In *Approaches to Conversion/Zero-Derivation*. Waxmann, Münster.

Hans Marchand. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach, 2nd edition*. C.H. Becksche Verlagsbuchhandlung, München.

David Pesetsky. 1995. *Zero syntax: Experiencers and cascades*. MIT Press, Cambridge.

Ingo Plag. 1999. *Morphological productivity: Structural constraints in English derivation*. Mouton de Gruyter, Berlin and New York.

Ingo Plag. 2003. *Word-Formation in English*. Cambridge University Press, Cambridge.

Laura Rimell. 2012. *Nominal Roots as Event Predicates in English Denominal Conversion Verbs*. Ph.D. thesis, New York University.

Enrico Santus, Alessandro Lenci, Qin Lu, and Sabine Schulte im Walde. 2014. Chasing hypernyms in vector spaces with entropy. In *Proceedings of EACL*, pages 38–42, Gothenburg, Sweden.

Petra Maria Vogel. 1996. *Wortarten und Wortartenwechsel. Zu Konversion und verwandten Erscheinungen im Deutschen und in anderen Sprachen*. Mouton de Gruyter, Berlin and New York.