VERB OR NOUN? WORD CONVERSION AND FREQUENCY EFFECTS IN ENGLISH[*]

Laura Teddiman
University of Alberta

1. Introduction

Conversion, the process through which one word is derived from another without overt marking, is a robust word-formation process in English (Plag 1999: 219; Plag 2003: 107). For example, the word *work* can be used as either a noun (*I have a lot of **work** to do*) or as a verb (*I **work** at the lab*), and is easily interpretable in both roles. What is not as clear is how speakers store such words or how they access them. For instance, while the category of a word might be clear in context, what happens when it is viewed by itself? This question forms the basis of the current study, and is explored through online experimentation and offline corpus data.

Three main proposals describing how categorically ambiguous words are encoded in the mental lexicon are lexical underspecification, zero-derivation, and storage. In *lexical underspecification*, the root exists in the mental lexicon, but is not specified for lexical category until the word is realized within a sentence (e.g., Barner & Bale, 2002). In other words, it is context that determines not only whether an ambiguous word is interpreted as a noun or a verb, but also whether it *is* a noun or a verb. In the case of *zero-derivation*, there is one lexical entry and it is stored as a member of a single lexical category. It must undergo affixation with a zero-morpheme to join a different lexical category (e.g., Harley, 2003). For example, *work* might be stored as a noun, but must be affixed with a zero-morpheme before it can be used as a verb. Finally, it is possible that each instance of an ambiguous word is stored, so that there is an entry in the mental lexicon for each word as a member of each lexical category (e.g., Don, 2004). For storage-based proposals, *work* would have two separate entries, one for the noun and one for the verb.

When ambiguous words are presented within a sentence context, they can be understood or are perceived as members of a single lexical category. Behaviour within a sentence context, however, cannot easily answer questions about the singular characteristics of ambiguous words, that is, whether speakers are aware of their lexical category outside of a sentence. How are words, particularly categorically ambiguous words, processed when presented in

[*] This research was supported by a Major Collaborative Research Initiative Grant (#412-2001-1009) from the Social Sciences and Humanities Research Council of Canada to Gary Libben (Director), Gonia Jarema, Eva Kehayia, Bruce Derwing, and Lori Buchanan (Co-investigators), and by a SSHRC Doctoral Fellowship. Any questions about the study presented here should be directed to Laura Teddiman (teddiman@ualberta.ca).

Actes du congrès annuel de l'Association canadienne de linguistique 2008.
Proceedings of the 2008 annual conference of the Canadian Linguistic Association.

isolation? This question is worth asking because it speaks to the underlying mental representation of words in the mind. This study explores the native speaker sensitivity to the lexical categories (noun, verb) of English monomorphemic words when they are presented in isolation, using a category decision task.

2. The Present Study

There were two main areas of interest in this study. The first focused on responses to categorically ambiguous English words gathered during an experimental procedure, and the second correlated this recorded online behaviour with category expectations calculated from a corpus of English. We first asked whether speakers are able to determine the lexical category of *any* word in isolation, and second, we asked whether the frequency with which each item occurs as a noun or a verb influenced participant choice. If native speakers of English do not encode information about lexical category without the use of a sentence frame, then performance in this task is expected to be at the level of chance. Likewise, if a sentence is required for the correct interpretation of the lexical category of an ambiguous word, then performance in this set of critical items should be random and at chance. If, however, speakers are sensitive to how often words occur as nouns or verbs, then there should be variation in responses based on the relative frequencies of ambiguous items.

2.1 Category Decision Task

In this experimental task, we asked whether participants were able to determine the lexical category of presented words in isolation. The task is similar to a standard lexical decision experiment, save that participants are asked to make a decision about word category instead of word status. Responses to unambiguous nouns, unambiguous verbs, and ambiguous words with preferences in either direction were of interest, although the categorically ambiguous words formed the target group.

Participants A total of 27 participants were involved in this study. Participants were primarily recruited from the undergraduate population at the University of Alberta and were compensated for their time.

Apparatus Experiments were scripted in PsyScope 1.2 and were presented on an Apple Macintosh computer.

Materials Stimuli consisted of 45 unambiguous nouns (e.g., *bird*), 45 unambiguous verbs (e.g., *earn*), and 34 categorically ambiguous words, some of which more commonly occur as nouns (e.g. *work*) and some of which are more commonly used as verbs (e.g. *walk*). Half of the ambiguous items had a preferred verbal reading and half have a preferred nominal reading. For each ambiguous word, its frequency as a noun was compared to its frequency as a verb. Items that occurred more often as nouns were considered to have a preference for nominal expression. For the purpose of analysis, answers were

considered correct if they agreed with the categorical preference of the target item.

Procedure Participants were seated a comfortable distance from a computer screen in a quiet environment. Items were displayed one at a time on the computer screen and participants were asked to decide whether the item was a noun or a verb as quickly and accurately as possible. They indicated their choice by pressing "z" for "noun" and "/" for "verb." The experiment took approximately five minutes. Error rates and reaction times were recorded.

Results & Discussion Participants were able to perform this task with a high degree of accuracy (Figure 1), and so were able to make determinations about the lexical category of given items without a supporting context. These results indicate that for unambiguous items such as *bird* and *earn*, participants do not require a larger sentence frame in order to categorize these items, and one might propose that they instead are relying on past language experience.

Responses to the ambiguous target items were more variable than they were to unambiguous words. Participants were significantly less accurate when asked to categorize ambiguous items as nouns or verbs, $F_2(3, 121) = 17.45$, $p<.001$ (Figure 1), but were still able to categorize ambiguous items with regard to their preferred lexical category 75 percent of the time. In other words, if an item was categorically ambiguous, but occurred more frequently as a noun than as a verb, then participants were more likely to decide that the item was a noun when asked.

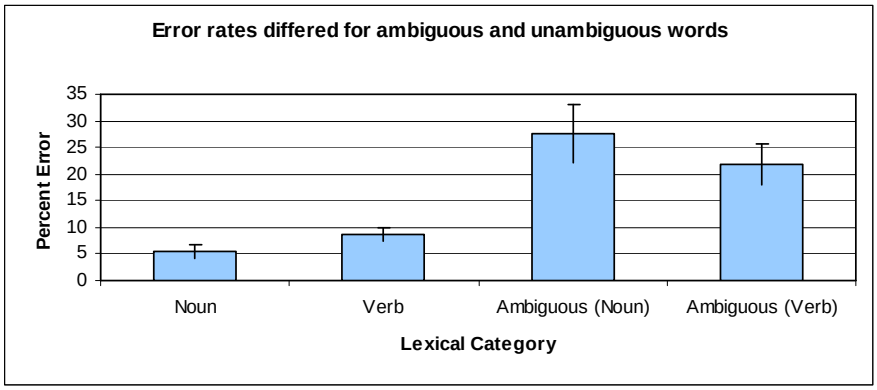

Figure 1. Participants were able to correctly indicate the lexical category of lexical items presented in isolation. Participants were less certain when categorizing ambiguous words, reflected in significantly higher error rates $F_2(3, 121) = 17.45$, $p<.001$.

Analysis of reaction times revealed no significant differences between any of the groups, including between ambiguous and unambiguous items. As reaction times may increase with increased difficulty, this lack of significance suggests that ambiguity did not increase the difficulty of categorization for these

items. It is possible that participants responded to the most accessible or most strongly associated word class when presented with ambiguous items, although it is unclear from these data whether both meanings were accessed during the task.

2.2 The Role of Relative Frequency

Target items used in the category decision task were categorically ambiguous words that could be interpreted as either a noun or a verb. As the preference for one category over the other was based on recorded frequency counts from the CELEX Lexical Database, it was possible to quantify the extent to which each target word preferred one lexical category over another. A simple measure was calculated to compare relative frequencies of nominal and verbal occurrence. The nominal frequency of each item was divided by the total observed frequency of the item, resulting in a number ranging from 0 – 1, with a score of "0" indicating a pure verb, and "1" indicating a pure noun. This was called the *CELEX Nominal Weight*. Similarly, an *Experimental Nominal Weight* was calculated to measure how strongly an experimental item was perceived to be a noun or a verb. This was done by dividing a count of the number of "noun" responses for a given item by the total responses to that item, and uses the same scale as the CELEX Nominal Weight.

When these two measures of nominal preference were compared, responses from the online category decision task were found to be strongly correlated with the offline measure for nominal weight, $r=.947$, $n=125$, $p<.000$ (Figure 2). Unambiguous words had CELEX Nominal Weights close to or equal to zero (verbs) or one (nouns), and in general, Experimental Nominal Weights for these items also tended to be close to zero or one, respectively. Ambiguous items were more or less preferred as a noun, as indicated by the CELEX Nominal Weight. The strong correlation between the CELEX Nominal Weight and the Experimental Nominal Weight suggests that as words became more ambiguous, there was more variability in participant classification, so that when the offline CELEX Nominal Weight nears the midpoint between noun and verb (0.5), participants were responding nearer to chance levels. Such items are equally likely to occur as nouns or verbs in a collection of texts, and participant behaviour correlates with this observation.

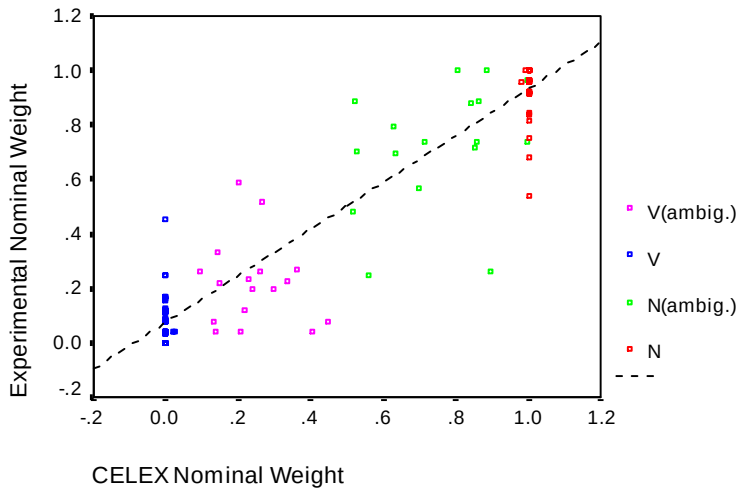

Figure 2. Relative frequencies of ambiguous words in CELEX are strongly correlated with category choice, r=.947, *n*=125, p<.000.

To summarize, when a word was more noun-like, it had a higher nominal weight in both the corpus and in the experimental responses. When a word was more verb-like, it had a lower nominal weight in the corpus and in the experiment. The highest level of ambiguity on the developed scales was at a weight of 0.5, representing items that are as likely to occur as nouns as they are to occur as verbs. As words became more ambiguous, as measured from the corpus, participants in the experiment also were as likely to decide that the item as a noun as they were to decide that it was a verb. Occurrence as a noun in the corpus highly correlated with a nominal decision in the category decision task. A corpus may be thought of as an indication of the prevalence of an item in the linguistic environment of an idealized speaker, and while not perfect, this correlation suggests the possibility that previous experience may be influencing current behaviour.

3. General Discussion

This study has provided evidence that speakers are more likely to categorize ambiguous words that have a strong preference for one lexical category over another as members of that preferred lexical category. Participants were able to correctly categorize most words without context, and the degree to which ambiguous items were judged to be nouns correlated with frequency data.

Creation of psycholinguistic stimuli can be a long and time-consuming process. The results presented might prove useful in the creation of stimuli where there are concerns about possible categorical ambiguities confounding results. Since it appears that ambiguous items can be interpreted by speakers as members of the category in which they most frequently occur, some degree of

conversion may be tolerated if research requires verbs or nouns to be presented in isolation. Results from the current data cannot determine the point at which an ambiguous item will be considered to be a member of one category with ninety-five percent accuracy, but do suggest that it might be possible to find that point with a larger dataset.

In the online experiment presented here, participants were asked to decide whether a word they saw on the screen was a noun or a verb. Some of the items were unambiguous, but the target items could be interpreted or perceived as a member of either category. Participants performed well on this task, with over 90% accuracy on unambiguous words, suggesting that they had access to the lexical category of the item, in some form, without depending on a sentence context. For these items, it appears that under the experimental conditions, the lexical category of the root was available to speakers. They made more errors on ambiguous items, but still performed above the level of chance. This result suggests that ambiguous items are more difficult to categorize without a supporting context, but that they may still have a 'main' reading that is related to the relative frequency with which it occurs as either a noun or a verb. It is possible that speakers can draw on their overall language experience with a word when it appears in isolation, and that this experience may influence how an item is stored in the lexicon.

References

Baayen, R.H., Piepenbrock, R. and Gulikers, L. 1995. *The CELEX Lexical Database* (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Barner, D. & Bale, A. 2002. No nouns, no verbs: psycholinguistic arguments in favor of lexical underspecificaiton. *Lingua*, 112, 771-791.

Don, J. 2004. Categories in the lexicon. *Linguistics*, 42(5), 931-956.

Harley, H. 2003. How do verbs get their names? Denominal verbs, Manner Incorporation, and the ontology of verb roots in English, Paper version of a 2001 talk at the Ben-Gurion University workshop on 'The syntax of aspect.'

Plag, I. 2003. *Word-Formation in English*. Cambridge, United Kingdom: Cambridge University Press.

Plag, I. 1999. *Morphological Productivity: Structural Constraints in English Derivation*. New York, United States of America: Mouton de Gruyter.