A Prosodic Analysis of Intervening Objects in English Phrasal Verbs Using the British National Corpus

Ethan Sherr-Ziarko, Jieun Kiaer

University of Oxford
essz.research@gmail.com, jieun.kiaer@orinst.ox.ac.uk

ABSTRACT

This paper describes a corpus-based study which investigates the connection between articulation rate and the phenomenon of heavy NP shift (i.e. pick **the book** up versus pick up **the book**), particularly as it relates to Hawkins' [7] theory of syntactic domain minimisation. Using data from the spoken BNC, we analyzed and compared the articulation rates of phrasal verb structures with and without intervening object NPs (e.g. pick **the book** up) as well as analyzing segment counts to determine which is more highly predictive of the intervener/non-intervener contrast.

We found both that speakers were using a 'squeezing' strategy to minimise the domains of phrasal verbs as the object NPs increased in length, and also that segment count did not serve as a useful predictor of the intervener/non-intervener contrast.

Keywords: Prosody, phrasal verbs, heavy NP shift, domain minimisation.

1. INTRODUCTION

Consider (1) from [16]:7:

(1) *Frequency asymmetry*:

Frequently observed:

(a) Pat picked up a very large mint-green hard cover book.

Hardly observed:
(b) Pat picked a very large mint-green hard cover book up.

The asymmetry observed in (1) is due to so-called Heavy NP (Noun Phrase) Shift accompanying a *phrasal verb*, a common syntactic structure in English wherein a verb is semantically paired with a particle [5],[15]. This phenomenon was observed first in [13] but this sort of weight-sensitive phenomenon has been observed since [3]. Though both sentences are grammatical, native speakers strongly prefer to locate 'heavy' NPs at the end of the sentence so that the verb and its accompanying particle may come together, as in (1a).

Previous work examining the phenomenon of heavy NP shift and the frequency asymmetry seen in (1) have largely been focused on the syntactic domain [7],[8],[11] in their efforts to explain when a phrasal verb object NP will be joined (as in 1a) or split (as in 1b, also referred to as an intervening object NP). In particular [7] and [11] investigated heavy NP shift using a number of corpora of primarily written English, and came to the conclusion that the word count of the object NP is the primary conditioning factor for the phenomenon. This idea was further formalized in [7],[8] which proposed the theory of *domain minimisation* to explain frequency asymmetry.

The theory of domain minimisation broadly states that the human processor prefers to minimize the size of the distance between related elements in a syntactic domain [7]:31, and argues that a structure like 1a, where the brain only has to process 3 immediate constituents to understand the phrase structure is much more efficient (and therefore preferred) than 1b, where 10 constituents must be processed. [7] also makes the argument that word-counting is the best methodology to use to predict whether a phrasal verb object NP can be an intervener.

While there is ample evidence in corpora of written English for the theories espoused in [7], there are some potential issues in spoken language. The most immediately apparent is that if speakers are concerned with the size of a syntactic domain, simply counting words may not be able to accurately capture the domain's length. It is perfectly possible for a single word to be longer than a group of three (e.g. 'a red book' versus 'accommodation'), and these words can be spoken at highly variable articulation rates by different speakers in different conditions.

In order to address this potential issue with existing investigations of heavy NP shift, this study makes use of data from the spoken portion of the British National Corpus [14] to determine whether prosody (particularly various aspects of articulation rate) plays a significant role in the intervener/non-intervener category split, and whether prosodic variables or word count are superior in predicting the category split. The hypothesis that this study will test is that there will be significant differences in the

prosody of intervening versus non-intervening object NPs, and that prosody will be significantly more useful in predicting the category split than word count in spoken English.

2. THE BNC STUDY

2.1. Data collection

Audio data for the study was obtained from the BNC, a large-scale corpus of written and spoken British English containing approximately 100 million words. The spoken portion of the BNC, which was the focus of this study, has been annotated and force-aligned using the Penn forced aligner, allowing for relatively straightforward browsing of the data. For the purposes of this project, the Lancaster BNCweb front-end was used to search the corpus.

Data was taken for 10 phrasal verbs in the corpus, in both their present and past tense forms where they differed. In choosing the phrasal verbs to analyse, firstly, their lemma frequency was considered; only phrasal verbs with a lemma frequency of greater than 100 parts-per-million were used [9], as lemma frequency has previously been found to have an effect on duration [6]. Only phrasal verbs made up of single-syllable words were chosen (i.e. *put up* is acceptable, but *frighten away* is not), and an effort was made to vary the preposition of the phrasal verb in order to minimise confounds related to examining an overly homogeneous set. Finally, in order to facilitate meaningful comparisons between the intervener and non-intervener categories, only phrasal verbs where the object NP could potentially appear as either were selected. The list of phrasal verbs examined in found in (2).

(2) (a) bring/brought up
(b) cut down
(c) fill/filled up
(d) give/gave back
(e) give/gave out
(f) pick/picked up
(g) put out
(h) take/took off
(i) take/took out
(j) turn/turned off

Only examples of phrasal verb structures where the object of the phrasal verb could potentially occur as either an intervener or non-intervener were selected. This meant that many single-word objects which would never occur as a non-intervener (i.e. pick *it* up) were excluded from the analysis. This resulted in 337 examples of interveners and 230 examples of non-interveners to be analysed. The experimental tokens are each recordings of spoken British English, containing a phrasal verb along with an object (either intervening or non-intervening) of 1-7 words, or 1-8 syllables.

Because the alignment of the BNC is not completely accurate, the chosen tokens were manually re-aligned, and their (canonical lexical) syllables manually counted in order to obtain accurate durational and articulation rate data. The method of syllable counting does leave something to be desired as no distinction is made between heavy and light syllables, but unfortunately segment-count meta data is not available for the BNC, and so a basic syllable count was the lowest level judged to be feasible. Pauses were included in the articulation rate data. The actual raw numbers were gathered using automated scripts based on Praat [4] text grids.

2.2. Statistical methods

The data in this study was analysed using two statistical tools. Firstly, data was analyzed using mixed effects regression models [2] in R [12], in order to allow us to include the random factor of the phrasal verb itself in the analysis as we have no particular hypothesis relating to the frequency of interveners versus non-interveners between the different phrasal verbs. Secondly, in order to test the relative usefulness of prosody versus segment counting to the intervener/non-intervener category distinction, a random forest analysis [10] (which allows for the evaluation of the predictive usefulness of each variable) was used.

2.3. Data overview and analysis

The variables examined in this study were the articulation rate (in both words/second and syllables/second) of both the object NPs and the full phrasal verb structures, the durations (in ms) of the object NPs and the full structures, and the segment counts in both words and syllables of the object NPs. Before comparing intervening and non-intervening objects using mixed models, there are a few points of interest regarding the data as a whole. Figure 1 shows the relationship between articulation rate and the syllable count of the object NP.

It is visually apparent from Figure 1 that speakers increase their articulation rate in a somewhat logarithmic fashion as syllable count increases. This pattern is present in both interveners and non-interveners, although the pattern is more pronounced and less stable in non-interveners. A modelling analysis as in (3) shows that this connection between syllable count and articulation rate is significant for both groups.

(3) **Model Structure**

Articulation Rate ~ Syllable Count + (1+Syllable Count|Phrasal Verb)

Intervener Model Results

Estimate: 0.32± 0.08, $X^2(1) = 11.45$, $p < 0.001$

Non-Intervener Model Results

Estimate: 0.14± 0.06, $X^2(1) = 5.43$, $p < 0.05$

This significant relationship could provide some evidence for the strategies underlying heavy NP shift, in that it appears that speakers are making an effort to 'squeeze' longer object NPs in order to shrink the domain of the phrasal verb. This observation can also be seen in another form when examining the *ratio* of the articulation rate of object NPs to the articulation rate of the full phrasal verb object, as shown in Figure 2.

Figure 1: Articulation rate (s/second) by number of syllables for both interveners and non-interveners.

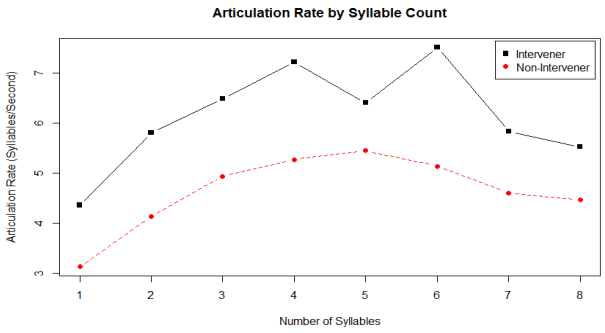
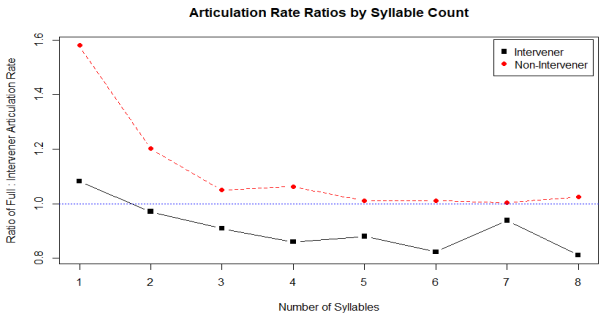


Figure 2: Ratio of articulation rate in object NPs to the articulation rate of the full phrasal verb object. The dotted line represents the point where the two rates are equal.



Based on Figure 2, it appears that as object NPs get longer, speakers make an effort to articulate the object NPs faster than the related phrasal verb. This pattern is notably stronger in intervening objects, and so the variable of articulation rate ratio was also included in the modelling analysis.

The modelling analysis, which compared variables of interest in the categories of phrasal verbs with interveners and those with non-interveners, showed that every prosodic variable was significantly different between the two categories. Most notably articulation rate was significantly higher in intervening objects than in non-interveners, which was potentially confounding the result showing duration as significantly shorter in intervening objects. The only variable tested that was not significantly different between the two categories was word count, which was somewhat surprising given previous results [11]. Results of all models are shown in Table 1.

Table 1: Results of all mixed models comparing the category of intervener and non-intervener NPs.

	Model Description			**Results**	
Variable	Estimate	Std. Error	t-value	$X^2(1)$	$p(>X^2)$
Full Rate (s/sec)	0.59	±0.13	4.67	14.72	<0.001
Obj. Rate (w/sec)	1.57	±0.17	9.19	38.79	<0.001
Full Rate (w/sec)	0.83	±0.12	7.15	28.12	<0.001
Obj. Rate (w/sec)	1.54	±0.14	11.13	37.14	<0.001
Full Duration (ms)	-310.49	±47.76	-6.5	20.19	<0.001
Obj. Duration (ms)	-359.95	±41.93	-8.58	25.03	<0.001
Rate Ratio (s/sec)	-0.15	±0.02	-8.4	26.46	<0.001
Rate Ratio (w/sec)	-0.19	±0.02	-10.52	33.31	<0.001
Words	-0.14	±0.07	-1.43	2.04	0.15
Syllables	-0.72	±0.14	-5.15	11.37	<0.001

2.4. Random forest analysis

In addition to the mixed modelling analysis, a random forest analysis was conducted to measure the predictive usefulness of the prosodic and segment-count variables. The random forest was trained on a random sample comprised of 75% of the full data set, and it was tested on the remaining 25%. Figure 3 shows a plot of the predictive importance of each variable (expressed as the decrease in accuracy if the variable is removed from the model), which was calculated as the model was built.

Figure 3: Predictive importance of each variable in the random forest.

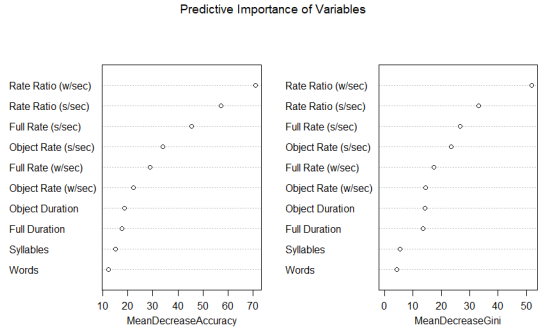


Figure 3 shows that articulation rate ratio is the most predictively useful variable, followed by

articulation rate in various forms. The durations are comparatively less useful, while word and syllable count are of very minimal use. This is borne out by the predictive results of the random forest: the model using all of the variables is able to correctly classify object NPs as interveners or non-interveners with ~79.5% accuracy, and if word and syllable count are removed from the model, accuracy decreases by less than 1% to ~78.8%. Although this does not invalidate the theory that counting words can allow us to predict the likelihood of a phrasal verb object being an intervener as word counting does have some predictive value, it does strongly support the hypothesis that prosody is much more predictively useful.

3. DISCUSSION

To review, this study tested the hypotheses that prosody plays a significant role in the contrast between the categories of intervening and non-intervening phrasal verb object NPs, and that prosodic variables would be more predictively useful than segment counting in distinguishing the two categories. These hypotheses were largely supported by the data, going against previous results [7],[11] which suggested that word-counting was a robust method for predicting heavy NP shift.

The first finding of interest is – as shown in Figure 1 – that speakers appear to be making an effort to shorten the duration of the syntactic domain of the phrasal verb by increasing their articulation rate as the object NP gets longer, and this pattern is particularly prevalent in cases of intervening objects (which are theorized in [7] to be more difficult to process as the size of the domain increases). This, in addition to the fact that speakers are increasing the rate of intervening object NPs to be greater than that of the rest of the phrasal verb object, can be taken as evidence for the theory of *domain minimisation* as it relates to heavy NP shift. It does appear that speakers are attempting to 'squeeze' more words into a limited syntactic domain as the object NP gets longer. This result does, however, suggest that there are issues for the word-counting method for analyzing heavy NP shift, as it is possible for speakers to use prosody to override potential limitations on the word counts of intervening objects.

This observation that prosody is important to the intervener/non-intervener category split was borne out by statistical analyses, which showed all tested prosodic variables as significantly different between the two categories, while differences in word count were not significant. Furthermore, a random forest analysis showed that prosodic variables were much more useful in predicting the category split than segment count. Taken together, the statistical analyses support the hypothesis that prosody is a critical part of the phenomenon of heavy NP shift, although it is most likely one of many combined conditioning factors (including syntactic considerations such as word count) that go in to determining whether an object NP will be an intervener.

The fact that word count did not appear statistically significant to the category distinction was somewhat surprising, but can likely be explained by the fact that most previous studies of heavy NP shift relied primarily on written corpora of English [11] and examined minimal spoken language. There are numerous well-attested differences in syntax and lemma selection between read and spoken language [1] which likely serve to explain the discrepancy between the current study and previous work.

4. CONCLUSION

Overall, the results of this study do in principle support Hawkins' [7] theory of *domain minimisation* as it relates to issues of locality in English phrasal verbs, while also demonstrating that prosody is a critical component to the same issues in spoken language. More generally, this study has shown that there are areas of potential overlap between syntax and phonetics, particularly when considering weight-sensitive phenomena such as heavy NP shift.

Although it is difficult to posit a direct causal relationship between prosody and heavy NP shift based on the evidence in this study, it is apparent that prosody should be considered in addition to purely syntactic criteria when researching phenomena that could in some way overlap with phonetics. Furthermore, it appears that the development of theories relating to weight-sensitive phenomena more generally should at minimum test for a significant relationship between the phenomenon of interest and potentially relevant prosodic variables.

Moving forward, future research could test for the salience of prosody to weight sensitive phenomenon more generally in both English, and other languages where such phenomena exist. A somewhat more controlled production study could also serve to test the relevance of prosodic factors such as f_0 or formants which could not be investigated in this study due to the sometimes poor quality of the BNC audio recordings.

5. REFERENCES

[1] Akinnaso, F. N. 1982. On the differences between spoken and written language. *Language and Speech*, 25(2), 97–125.

[2] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1), 1–48.

[3] Behaghel, O. 1909/10. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Insogermnische Forschungen*, 25,110–142.

[4] Boersma, P., Weenik, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/

[5] Fraser, B. 1976. *The verb-particle combination in English*. New York: Academic Press.

[6] Gahl, S. 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3), 474-496.

[7] Hawkins, J. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

[8] Hawkins, J. 2011. Discontinuous Dependencies in Corpus Selections: Particle Verbs and their Relevance for Current Issues in Language Processing. In J. Arnold & E. Bender (eds.), *Readings in Cognitive Science*. Stanford, California: CSLI Publications, 269–291.

[9] Leech, G., Rayson, P. 2014. W*ord frequencies in written and spoken English: Based on the British National Corpus*. Abingdon: Routledge

[10] Liaw, A., Wiener, M. 2002. Classification and Regression by randomForest. *R News* 2(3), 18–22.

[11] Lohse, B., Hawkins, J., Wasow, T. 2004. Domain minimization in English verb-particle constructions. *Language*, 80, 238–261.

[12] R Core Team. 2018. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria.

[13] Ross, J. R. 1967. *Constraints on variables in syntax*. PhD Dissertation, MIT.

[14] *The British National Corpus*, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

[15] Thim, S. 2012. *Phrasal verbs: the English verb-particle construction and its history*. Berlin: De Gruyter Mouton.

[16] Wasow, T. 2002. *Post-verbal Behaviours*, Stanford: CSLI.