Allomorphs of French *de* in coordination
a reproducible study[*]

How much influence can phonology have on a surface syntactic form? An answer at one extreme is that morpheme choice and order are fully determined before the phonological content of those morphemes is accessed; all the phonological module can do is fill in their sound content—that is, phonology is late. At the other extreme—early phonology—phonological considerations have the same qualitative ability as syntactic considerations to influence morpheme choice and order, and even syntactic structure.

In practice, both theoretical and processing models have adopted a variety of intermediate positions. For example, Embick and Noyer (2001), working within Distributed Morphology, adopt an essentially late-phonology stance, but with Local Dislocation, a restricted mechanism by which morpheme order can change in response to phonological information. By contrast, Anttila, Adams, and Speriosu (2010), building on Bresnan et al.'s (2007) work on the English dative alternation, propose an Optimality Theoretic (Prince & Smolensky 1993) grammar in which the syntactic constraint "The goal NP forms an XP with its head" competes directly with phonological constraints such as "No stress clashes within a phonological phrase" (p. 962). While the syntactic constraint is one of the highest-ranked in the grammar, it is not qualitatively prior to phonological constraints. This short article will not attempt a full history of this debate. Pullum and Zwicky (1988 and others) argue for late phonology, and exclude "preferences and tendencies" (p. 274) from being possible counterexamples. Within the field of phonology, however, preferences and tendencies have been treated as qualitatively similar to categorical grammaticality; it has been argued that the same factors that can render a form phonologically ungrammatical in one language or context can have a gradient effect in another, or even that apparent categoricalness is merely an extreme form of variation (e.g., Pierrehumbert 1994, Boersma 1998, Cohn 2006, Hayes & Londe 2006 for a variety of perspectives). Shih 2014 provides a recent survey of the issue, including new cases from English variation that support early phonology.

The goals of this paper are empirical and methodological. On the empirical side, it argues that phonological information must be allowed to influence, albeit probabilistically, whether French *de* 'of' appears once or twice within a coordinated structure. We will consider whether the French case must reflect phonological influence on choice of syntactic structure, or could be accounted for by giving the phonology the power to delete the second *de*. This case thus adds to the body of data whose implications for the syntax-phonology interface can be debated.

On the methodological side, this paper strives to meet the evolving standards of reproducible research (Stodden, Leisch & Peng 2014 is a recent handbook). *Replication* of a study typically begins from data collection, such as running a new experiment, or gathering a new corpus. If another researcher can perform these steps and obtain similar results, confidence in the original finding increases. The term *reproducible research*, on the other hand, refers to research that makes it easy for a reader to start from the same raw data as the original study, and re-generate and scrutinize the original analysis—that is, only part of the original research pipeline is repeated. Reproduction is useful as a substitute for replication when it would be infeasible to run a new experiment, gather a new corpus, etc.: with much less effort, many of the benefits of replication can still be obtained. This paper begins from a publicly available corpus, and the accompanying web materials, archived at TROLLing (The Tromsø Repository of Language and Linguistics, opendata.uit.no/dvn/dv/trolling), include all scripts used for data processing and analysis, with ample documentation, so that the interested reader can reproduce all results, statistical tests, and figures; make changes to the assumptions and methods; and look for errors, with only one departure from the script for some hand-coding, which the reader can either use as-is or perform independently.

1 Background: French *de* and coordination

The French preposition *de*, meaning roughly 'of', has two properties crucial to this study: it can optionally take wide scope in coordination, and it has different allomorphs depending on whether the following word begins with a consonant or a vowel. This section describes each of these properties in turn.

French *de* can optionally take wide scope over a coordinated structure. Examples are given in (1), with *de* in boldface, brackets marking selected constituents, and the two coordinands underlined.

(1) a. morceaux [**de** tomates] et [**de** carottes]
        pieces   of tomatoes and of carrots

    or

    b. morceaux **de** [tomates   et   carottes]
        pieces   of   tomatoes and carrots

        'pieces of tomatoes and (of) carrots'

Miller 1992 appears to be the earliest formal treatment of wide scope versus repetition for French *de* (among other function words). (The terms *wide scope* and *reduction* will be used interchangeably below to describe data like (1)b; *repetition* will describe cases like (1)a.)

This study examines whether phonological properties of the words involved affect the frequency of wide scope. I assume the syntactic bracketings suggested in (1): for repetition, constituents [de X] and [de Y] are coordinated; for wide scope/reduction, *de* is outside a constituent

[X & Y]. The conclusion will consider (and reject) the alternate possibility that *de* never actually takes wide scope, and (1)b has the structure [[de tomates] et [~~de~~ carottes]].

There are some non-phonological restrictions on wide scope. Miller cites claims from a classical and still popular descriptive grammar (Grevisse 1980) about the conditions under which *de* can take wide scope, but he does not agree with all of them. Miller concludes that repetition is the unmarked case, occurring more frequently than wide scope and imposing no special semantic restrictions, whereas wide scope requires a group reading, at least in some cases (p. 162).

Tseng and colleagues, in a series of articles (Tseng 2003, Abeillé et al. 2004, Abeillé et al. 2005, Abeillé et al. 2006) make several further observations. They divide *de* into what they call oblique and non-oblique versions, where only the oblique *de* can take wide scope. Non-oblique *de* occurs with partitive *the*+N', bare N', and VP; oblique *de* occurs with NPs, PPs, APs, AdvPs, or lexical Vs (Abeillé et al. 2006, p. 151):[1]

(2) Non-oblique—repetition required
    a. *the*+N': il faut [**de** la farine] et [*(**de**) la levure]
                it needs of the flour   and   of the yeast
                'you need flour and yeast'

    b. bare N': beacoup [**de** pain] et   [*(**de**) vin]
                much   of bread and   of wine
                'a lot of bread and wine'

    c. VP: Je rêve [**de** lire   ce livre] et [*(**de**) l' expliquer à mon fils]
                I dream of to.read this book and   of it to.explain to my son
                'I dream of reading this book and explaining it to my son'

(3) Oblique—wide scope allowed
    a. NP: J' ai   besoin **de** [[cette farine] et [cette levure]]
                I have need of   this   flour   and this yeast
                'I need this flour and this yeast'

    b. PP: Il revient **de** [[chez   Paul] ou [chez   Marie]]
                he returns from home.of Paul or home.of Marie
                'He's coming back from Paul or Marie's'

    c. AP: quelqu'un **de** bon en maths et (**de**) fort   en gym
                someone of good in math and of strong in gym
                'someone good at math and strong in gym'

    d. V: Je rêve **de** [lire   et expliquer] ce livre à mon fils

---

[1] (2) is adapted from Abeillé et al. 2006's (58)-(60), p. 151. (3) is adapted from Abeillé et al. 2006's (61)-(64), p. 151, and footnote 8, p. 160. Brackets and boldface are mine.

I dream of to.read and to.explain this book to my   son
'I dream of reading and explaining this book to my son'

This study looks only at the coordination of single words. Unfortunately, it is not possible to tell whether a single N is itself an NP, or whether a single V is a VP, so inevitably the data contains some oblique instances (where repetition of *de* is claimed to be obligatory), mixed in with the non-oblique instances (where the variation of interest would lie).

Some of the phonological hypotheses to be tested below arise from the well-known fact that *de* has two allomorphs: *de* [də] is used before a consonant-initial word, and *d'* [d] before a vowel-initial word. Certain complications arise in glide-initial words and some words that are spelled with initial silent *h*, but those cases will be excluded from the data here. Some examples, from Abeillé et al. 2004 p. 16:

(4) CC-initial: de clients   [**də**kliã]   'of clients'
    C-initial: de livres   [**də**livʁ]   'of books'
    V-initial: d'enfants   [**d**ãfã]   'of children'

2 Hypotheses[2]

For phonological information to influence the choice between repetition and wide scope, there must be phonological differences among coordinands that cause omitting *de* to have different consequences. That is, reduction of *de* has to produce a better result for some words than for others, because of phonological properties of those words. The sense in which the result is worse (or better) could be phonological (relying on known properties of French phonology or of phonological typology generally), could involve processing, or could even involve the syntax-phonology interface.

Some of those consequences arise from the *de/d'* allomorphy. The following four cases arise for coordinating words *X* and *Y*, where C stands for 'consonant-initial word', V stands for 'vowel-initial word', and & stands for *et* 'and' or *ou* 'or':

(5) *type*   *example*
    C&C: de X et/ou (de) Y   de biens et (de) services   'of goods and (of) services'
    C&V: de X et/ou (d')Y   de juillet et (d')août   'of July and (of) August'
    V&C: d' X et/ou (de) Y   d'aller et (de) venir   'of going and (of) coming'
    V&V: d' X et/ou (d')Y   d'Adam et (d')Eve   'of Adam and (of) Eve'

The allomorphy suggests the following hypotheses:

---

[2] Thanks to Stephanie Shih for suggesting the hypotheses of Hiatus avoidance/Lapse avoidance and Repetition avoidance, and to Bruce Hayes for suggesting expanding Repetition avoidance to include [t].

(6) *Hiatus avoidance:*[3] *Et* [e] and *ou* [u] both end with vowels, so reduction when the second conjunct begins with a vowel causes a V-V sequence (e.g., *et Eve* [e ɛv] vs. *et d'Eve* [e dɛv]). Assuming this is penalized, reduction is less likely before V. (E.g., Prince & Smolensky's (1993) ONSET; for the role of hiatus avoidance in French specifically, see Tranel 1996, among many others)
　　　→ more reduction in X&C than in X&V

(7) *Dependent* d'*:* Although the allomorph *de* has schwa as its sole vowel, and therefore does not qualify as a phonological word of French, it can nonetheless stand alone as the citation form of the morpheme, whereas *d'* cannot—it can only be uttered when cliticized to another word. Assuming that *de* is a more independent unit than *d'*, *de* is more likely to take wide scope. That is, there is some penalty incurred by the structure *[d' [X & Y]]*, where *d'* forms a prosodic but not a syntactic unit with *X*.
　　　→ more reduction in C&X than in V&X (where X is either C or V)

(8) *Parallelism:* As mentioned above, we will consider below the possibility that wide scope is actually the result of deleting the second *de*. Deletion in conjunction is normally of a repeated element, so it should be more likely when the two *de*s, if both present, would appear in the same allomorph. That is, reduction is more likely when the conjuncts would take the same allomorph.
　　　→ more reduction in C&C and V&V than in C&V and V&C

All three hypotheses require phonological information about the two conjuncts to, at least sometimes, be available before the final decision about reduction-versus-repetition is made
　　　A final hypothesis requires not only prosodic information (C-initial or V-initial, syllable count) but even segmental information to be available:

(9) *Repetition avoidance:* Assuming that sequences $C_i \partial C_i$ are penalized, *de* is more likely to be omitted if the following word begins with [d], thus avoiding a sequence [də d...]. (E.g., Yip 1995, Raffelsiefen's 1999 ONS$_i$ONS$_i$ constraint, Löfstedt 2010)
　　　→ more reduction in X&d... than elsewhere

We will see, however, that this last hypothesis is not supported.

3　Methods

The methods section of this paper is in two parts: the brief version here, and the full details given in the much-longer HTML file that accompanies the paper. The HTML file gives a full narrative of the method, from downloading the initial corpus files to the final regression models. The narrative is interspersed with extensively commented code and the output of that code, including

---

[3] *Lapse avoidance* makes the same prediction*: Et/ou de* is a sequence of two unstressed syllables, but *et/ou d'* adds only one unstressed syllable between the two conjuncts. Assuming that stress lapses are penalized, and the more so the longe they are, reduction is more likely when the allomorph would be *de*. (E.g., Green & Kenstowicz's (1995) LAPSE).

5

figures. The HTML file is more than a lab notebook, though, because it is generated by an .Rmd ("R markdown") file, which contains the narrative and the R code (R Core Team 2014); the .Rmd file runs the R code and generates the HTML file using the knitr package (Xie 2014a; Xie 2014b). This makes it easy to alter and re-run the analysis. Also downloadable are a Python script and helper file that the R code calls to run part of the analysis, and various intermediate versions of the data files for readers who wish to skip steps.

Data come from the Google n-grams corpus (Michel et al. 2011, Lin et al. 2012) for French. The raw data (not available to the public and not used here) are scans of books published in French, digitized through optical character recognition. The downloadable n-gram corpus used here compiles word sequences (such as *d' aller et de venir*) and reports, for every publication year, how many tokens of that sequence occurred.

As the accompanying HTML file details, relevant 4- and 5-grams were extracted. The data used come only from the year 1900 and later, exclude conjuncts that begin with glides or orthographic *h*, exclude numerals, and include only conjuncts that are semantically parallel (e.g., two places, two materials, two occupations).

To test the hypotheses outlined above, each word1/conjunction/word2 combination's *reduction rate* is calculated. This will be used as the dependent variable in the plots below. For example, *d'aller et de venir* occurs 7,616 times and *d'aller et venir* occurs 12,696 times, so the reduction rate for *aller/et/venir* is 0.63. The *reduction count* used in the Poisson regression model below would be 12,696 in this case, and the *exposure* would be 7,616 + 12,696.

Some independent variables used below were calculated from the n-gram dataset: total frequency of word1/conjunction/word2, initial sound types of word1 and word2. Others were looked up in the Lexique 3.80 database (Matos et al. 2001, New et al. 2004): word1 and word2 frequency, syllable count, and part of speech.

After various types of data cleaning, the final data set includes 8,332 *word1/conjunction/word2* triples, representing 10,061,917 tokens.

4   Results

4.1   Descriptive statistics

The overall reduction rate is strongly skewed, with most items tending not to undergo reduction. This is in line with Miller's 1992 treatment of repetition as the unmarked case. The following histogram illustrates the distribution of reduction rates:
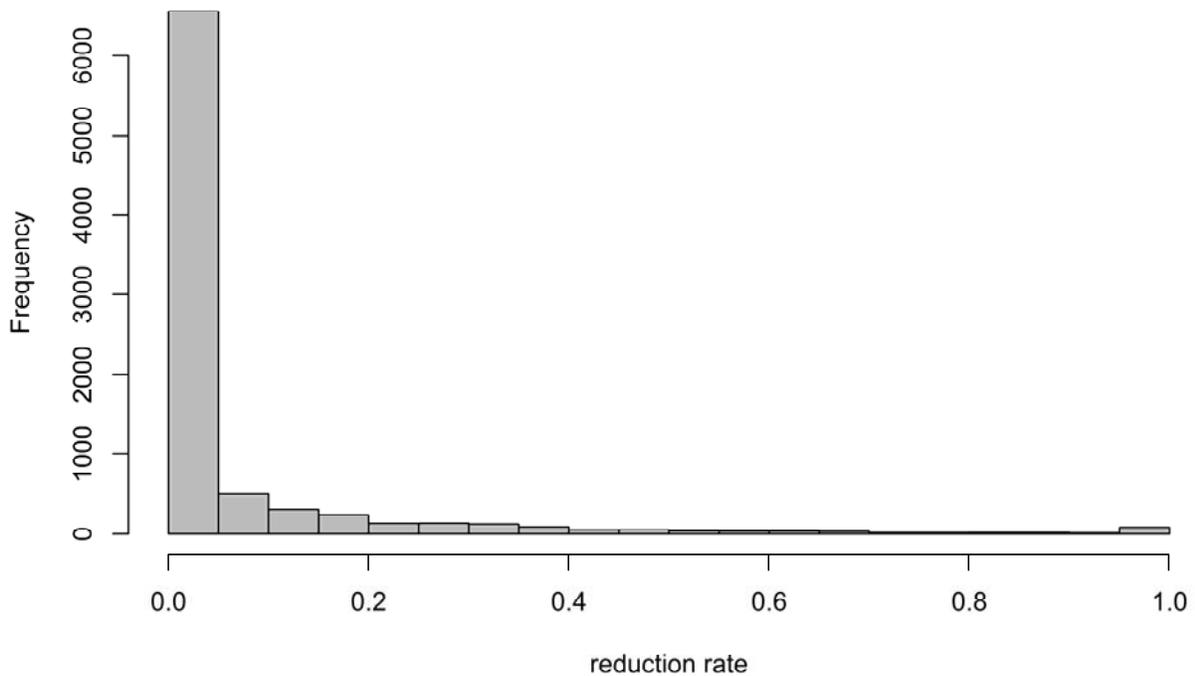
Figure 1: Histogram of *de*-reduction rates in French *word1/conjunction/word2* tuples

Note that there is a poor-get-poorer artefact here: the Google n-grams corpus includes only items that occur in at least 40 books (over the whole timespan of the corpus; the number of occurrences used here, since 1900, may be smaller). Suppose that *de X et de Y* occurs 270 times (with hits coming from 40 or more books) and reduced *de X et Y* occurs only 30 times in the whole corpus, falling below the threshold. The true reduction rate is 10%, but in the Google n-grams corpus it will appear to be 0%. The real distribution must be slightly less skewed than what we see in Figure 1.

To test the hypotheses concerning the allomorphs *de* and *d'*, the following bar plot breaks down mean reduction rates by the conjuncts' initial consonant type, separately for *et* and *ou*. Each category is coded as *word1 type, conjunction, word2 type*. For example, *C et V* means that the first conjunct is a consonant-initial word, the second is vowel-initial, and they are coordinated with *et*.
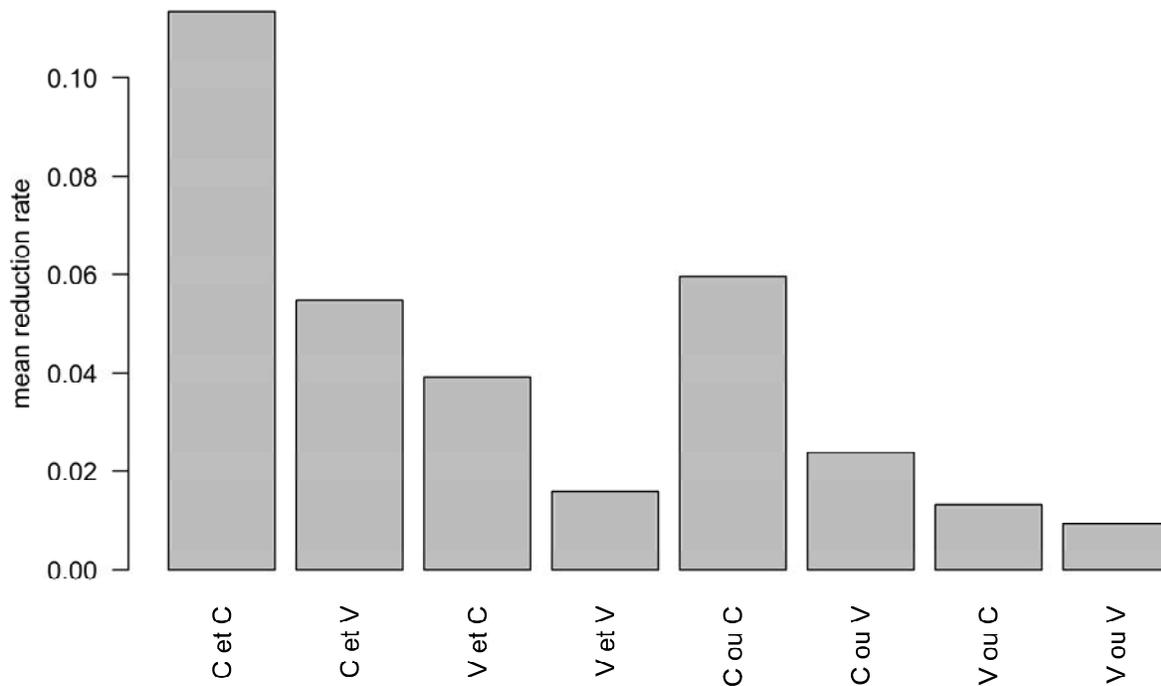
Figure 2: Mean *de*-reduction rates broken down by C- vs. V-initial

Figure 2 shows that, within both *et* and *ou* (which differ in their overall reduction rates), there is more reduction if word1 begins with C than if it begins with V (consistent with the Dependent *d'* hypothesis), and more if word2 begins with C than V (consistent with Hiatus avoidance/Lapse avoidance). It is hard to see from the plot whether Parallelism is supported—e.g, whether there is more reduction for *V&V* than would be expected from the overall rate of reduction for V in each position.

Reassuringly, the pattern holds within each part of speech, as plotted in Figure 3. Omitting adverbs, for which data are sparse, we can see the same pattern within adjectives (ADJ), nouns (NOM), and verbs (VER), though the overall rates differ by part of speech. This makes it unlikely that the effects are merely driven by particular words or constructions.

The lower rates seen for nouns and verbs could be because of the mixture of N' and NP, and V and VP mentioned in section 1.
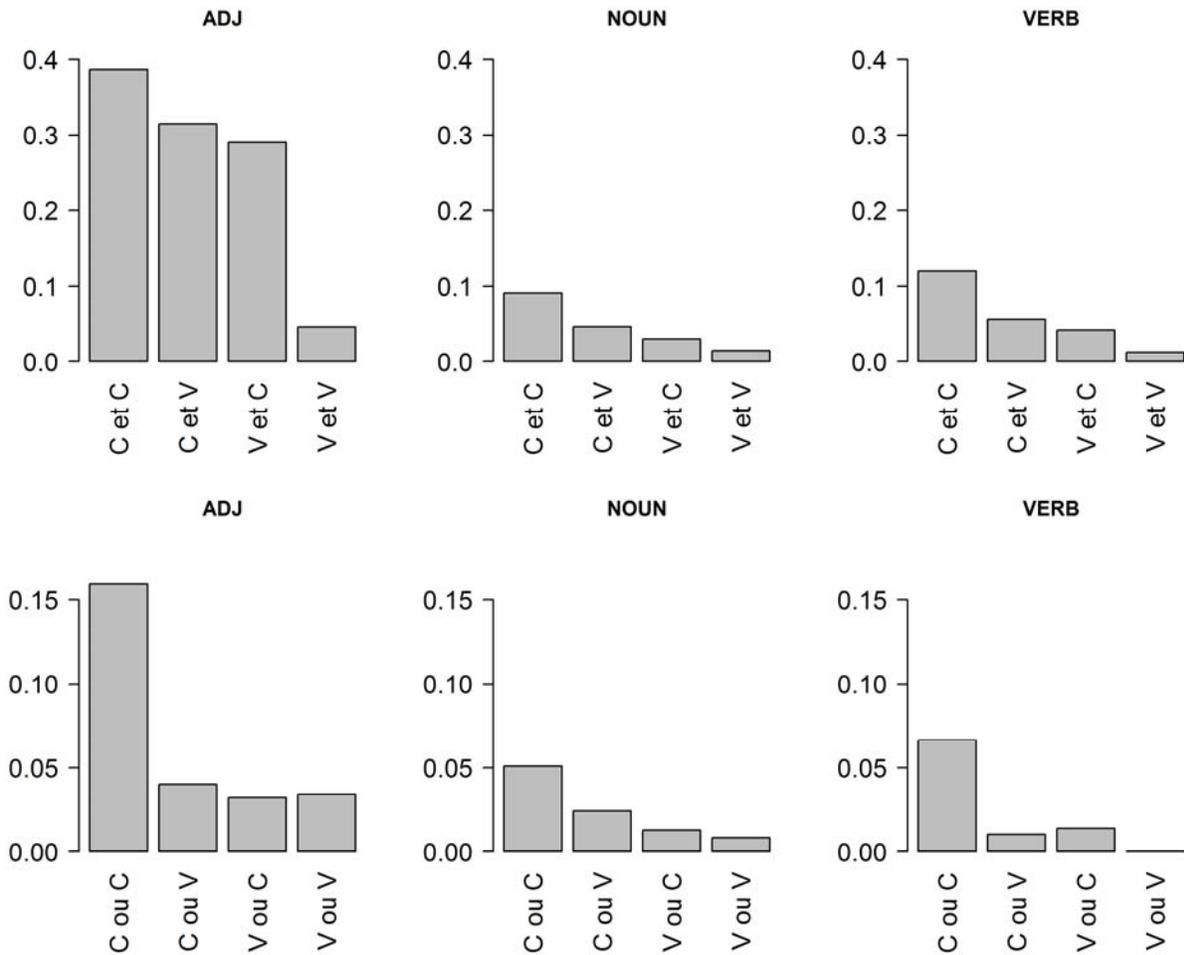
Figure 3: Mean *de*-reduction rates broken down by part of speech

Finally, one might wonder whether these phonological effects are frozen in high-frequency items, or apply productively to new combinations. To test this, word1/coord/word2 triples are divided into four frequency quartiles (frequency of 8-128, 129-474, 475-1,237, and 1,238-110,818), and the data are plotted separately for each quartile. The only place where the pattern fails to hold is for *V et C* in the lowest frequency quartile. Overall reduction rates are lower in the lower frequency quartiles (as confirmed in the regression model below). It seems plausible that the more frequent a combination, the more likely the shorter form (e.g., Bybee 2001 on reduction of all kinds in frequent items). We can also observe that the *et/ou* difference is driven by the lowest frequency quartile.
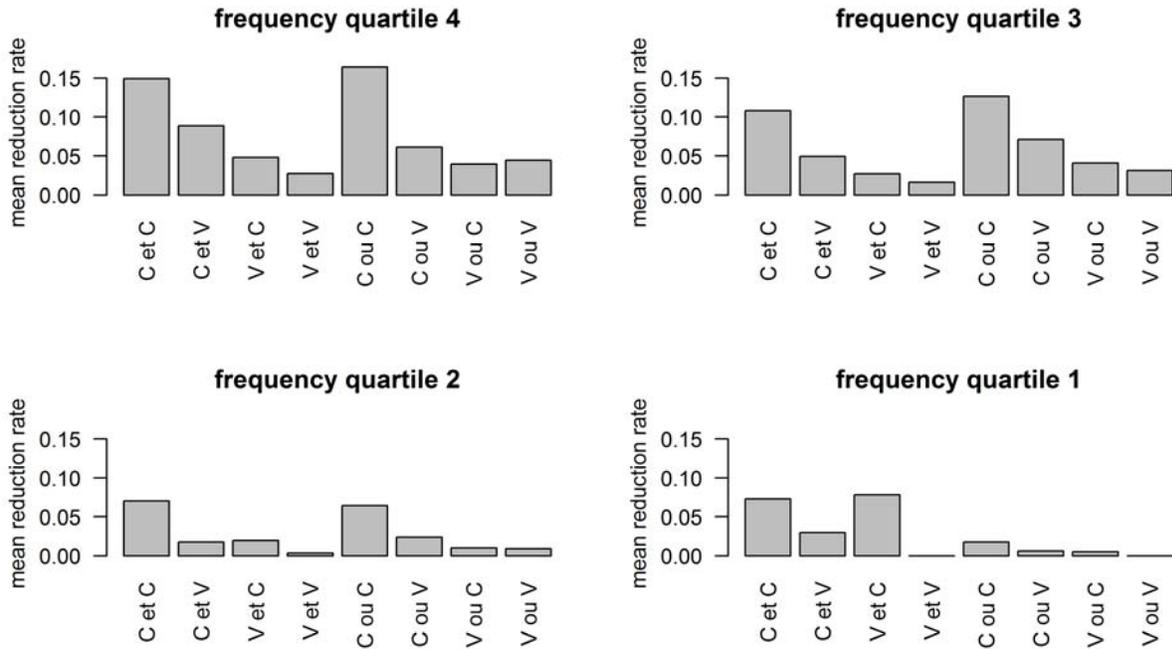
Figure 4: Mean *de*-reduction rates broken down by part of speech

4.2   Regression models

Because reduction rates are so strongly skewed, with most tuples having a value near zero, Poisson regression is used. (See HTML file for full regression details and results.)

Each observation is one word1/conjunction/word2 combination. The dependent variable is the count of reduced tokens, and the log total number of tokens (exposure) is used as an independent variable.[4] A random intercept is included for word2—that is, the model allows that each word2 could have its own idiosyncratic baseline rate of reduction (there is no random intercept for word1 because of convergence problems; it seems less plausible that idiosyncractic properties of word1 would have a large influence). The fixed effects are as follows:

non-phonological factors
- log frequency of the word1/conjunction/word2 combination
- log frequencies (from Lexique) of word1 and word2
- part of speech: baseline is adjective, with adverb, noun, and verb being contrasted

---

[4] The log exposure would be used as an "offset" if the rate of reduction were constant across values of exposure, but as we have seen, reduction rates are higher among higher-frequency items.

de *allomorph factors, in interaction with conjunction*
- word1 type (initial sound is C or V) and word2 type
- whether word1 and word2's initial sound types match[5]
- conjunction: *et* or *ou*[6]

segmental factor
- whether word2 starts with [d]

The model that was fitted is given below. Factors have been renamed for easier reading, but bear in mind all factors except part of speech and (log) total frequency were centered and, if numerical, standardized (by dividing by two standard deviations). Phonological main effects are bolded. The model was not able to converge using the *glmer* function of the lme4 package (Bates et al. 2014), so instead the *MCMCglmm* function of the MCMCglmm package (Hadfield 2010) was used to estimate coefficients, confidence intervals for coefficients, and a *p* value:

(10) Poisson regression model (fixed effects)

```
                          post.mean l-95% CI u-95% CI eff.samp  pMCMC
(Intercept)                -17.7279 -18.7216 -16.7737    93.68 <0.001 ***
log combination frequency    2.6709   2.5464   2.7737   112.11 <0.001 ***
log word1 frequency          0.0549  -0.2027   0.3104   605.23  0.664
log word2 frequency         -1.2920  -1.6523  -0.8781   586.41 <0.001 ***
POS = adv. (vs. adj.)       -0.8498  -2.2353   0.2852   633.15  0.198
POS = noun (vs. adj.)       -0.8954  -1.3151  -0.4978   806.91 <0.001 ***
POS = verb (vs. adj.)        1.5921   0.7178   2.4504   684.17 <0.001 **
w1 is V-initial (vs. C-init.) -2.2124  -2.5491  -1.8514   246.40 <0.001 ***
w2 is V-initial (vs. C-init.) -1.7439  -2.2163  -1.1508   342.31 <0.001 ***
w1 type (C or V) = w2 type    0.2865   0.0065   0.6134   293.61  0.056 .
conjunction = ou            -0.7711  -1.0476  -0.5205   356.52 <0.001 ***
word2 begins with d         -0.0584  -0.7841   0.8069   346.72  0.892
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the regression model fits logs of reduction counts, the results are interpreted differently than in other types of regression. For example, when comparing two w1/conj/w2 combinations, if the log frequency of word1 is greater in one of the combinations by 1 standardized unit (i.e., 2 standard deviations), then, all else being equal, the difference in the logs of their reduced-form counts is predicted to increase by 0.0549. As in other types of regression, a positive coefficient estimate means a factor favors *de*-reduction and a negative value means it disfavors it.

Considering non-phonological predictors first, combinations with higher frequency are subject to more reduction, as we have seen; the frequency of word2 has a smaller effect, in the other direction (a more-frequent word2 has less reduction). As we saw in the plots above, *ou* has a lower rate of reduction than *et*.

Initial regression models found more reduction when word2 begins with [d], but using the final Poisson approach, this factor made no significant contribution.

---

[5] This was thought to be easier to interpret that simply using the interaction of word1 type and word2 type.
[6] Originally, the model included the interaction of conjunction with the two allomorph factors, but the interactions' contributions were far from significant.

As for allomorphs of *de*, as we saw in the plots above, reduction is less frequent if either word1 or word2 begins with a vowel. There is also a small effect of matching: if word1 and word2 have the same status (C&C or V&V), then reduction is slightly more frequent than would be expected from the statuses of word1 and word2 independently.

5   Summary, discussion, and conclusion

The following table summarizes the results obtained for the hypotheses put forth in section 2:

hypothesis	*result*
Hiatus avoidance	Strongly supported: Less reduction when word2 is V-initial, and thus would produce hiatus if *d'* is omitted
Dependent d'	Strongly supported: Less reduction when word1 is V-initial, and thus would take *d'*
Parallelism	Supported (significant, but small coefficient): Slightly more reduction when word1 and word2 would take the same allomorph of *de*.
Repetition avoidance	Not supported: No increase in reduction when word2 starts with [d], so that retaining *de* would yield [dəd] sequence

Table 1: Summary of hypotheses

It is hard to see how the French data could be consistent with a theory in which phonological information has no influence at all on the choice to repeat or not repeat *de*. Whether the data are consistent with a late, limited mechanism or require an earlier or more active role for phonology will depend on the syntactic analysis of wide-scope *de*.

One possibility is that there really are two syntactic structures, as given in (1): [de X] & [de Y] and [de [X&Y]]. In that case, the syntax could produce multiple candidates, which are then operated on by the phonological component, and the final choice is made at least partly on grounds of phonological well-formedness. Or, still assuming two different syntactic structures, phonological information could simply be available while syntactic decisions are being made. From a processing point of view, access to phonological information would most likely unfold gradually, so that it couldn't be depended on for a categorical syntactic rule, but on occasions when phonological information was available before the speaker had committed to a syntactic structure, it would have a say in the decision (see Wagner 2011 on the idea that allomorphy is constrained by availability of phonological information during production planning). In both proposals, phonology is early.

Another possibility is that the syntax outputs a structure [[de X] & [de Y]], without reference to phonological information, but the phonology can delete of the second *de*, presumably without subsequent syntactic reorganization. That is, the syntactic structure of reduced forms

would be [[de X] & [de Y]], and the first *de* would not in fact take wide scope. This proposal would maintain that phonology is late.

Several athors have weighed in on the question of coordination vs. deletion when it comes to coordination in compounds or affixed words, as in English *half-(siblings) or step-siblings*, *step-sons and (step)-daughters.* Booij (1985), working mainly with Dutch and German data, argues that these structures arise through deletion rather than coordination. His reasons include: it is possible to "coordinate" non-constituents, as in *the difference between **a third-** and **a sixth-**grader*; it is possible to "coordinate" items of different syntactic category, as in Dutch ***wis-** en **sterren-**kunde* 'math and astronomy', literally 'sure- and stars-knowledge'; and it is not only conjunctions that allow this kind of deletion, but also other constructions, including prepositions, as in *how to distinguish **neuro-** from **psycho**linguistic claims* (Chaves 2008, p. 266). Chaves (2008) adds to Booij's arguments the observation that utterances like the following are possible: ***Pre-** and **post-**revolutionary France were very different from each other* (p. 264). Two different versions of France are in question (pre-revolutionary France and post-revolutionary France), not a single version of France that is *[pre- and post-]revolutionary*. Artstein 2005, by contrast, argues for coordination, in part because of a claimed semantic difference between *Mary and Bill are orthodontists and periodontists* (they must each practice both specialties) and *Mary and Bill are **ortho-** and **perio**dontists* (Mary and Bill may each have a different dental specialty). Artstein also points out that it seems strange for a phonological deletion rule to apply only next to conjunctions (and some prepositions): why not **My step-~~brother~~ likes his half-brother*? More generally, allowing the phonology to have a deletion rule that is sensitive to repetition in coordinated structure pushes at the boundary of the types of syntactic information that should be available to the phonology.

But the main challenge for treating *de* reduction as like phonological deletion is that it would be a rather different rule from within-word deletions. Usually in languages that allow deletion or coordination of word parts, both the remnant and the deleted part have to be at least a phonological word (or, in Artstein's 2005 analysis, a separate foot). Addressing the size of the remnant, Abeillé (2006) says,"Coordination of prefixes is marginally available in French, but only for prefixes with some phonological autonomy" (p. 3), and gives these examples: ***sur-** ou **sous-**évaluer* 'over- or under-evaluate', but *****re-** ou **dé-**faire* 're- or un-do'; [ʁə] lacks full vowel and is presumably sub-minimal as a French word. Addressing the size of the deleted part, Booij 1985 contrasts examples such as ***regel**-ordening en –toepassing* 'rule-ordering and –application' with *****ge**-hijg en –puf* 'gasping and puffing', where *ge-* is not a foot or a phonological word. In French *de X & Y*, the remnant of deletion would be *Y*, which poses no prosodic problem, because Y will always be at least a full phonological word. The deleted portion, however, is the subminimal clitic *de* or *d'*. Thus, *de* omission would be different from the within-word coordination/deletions discussed by Booij (1985), Artstein (2005), Chaves (2008), and Abeillé (2006). It is possible that

within-word coordination really is deletion, but that *de* omission instead reflects a different, wide-scope syntactic structure.[7]

In summary, the choice between French *de X & de Y* and *de X & Y* depends on phonological properties of both X and Y, including their influence on what allomorph of *de* is used. This could mean that the choice is itself phonological one (whether to phonologically delete the second *de*), but deletion would have to be phonologically different from deletion that has been argued for within words. Or, it could mean that phonological information is available early enough to influence syntactic structure.

References

Abeillé, Anne. 2006. In defense of lexical coordination.

Abeillé, Anne, Olivier Bonami, Danièle Godard & Jesse Tseng. 2004. The syntax of French de-N' phrases. In S Müller (ed.), *HPSG 2004: Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, 6–26. CSLI Publications.

Abeillé, Anne, Olivier Bonami, Danièle Godard & Jesse Tseng. 2005. Les syntagmes nominaux français de la forme de-N'. *Travaux de linguistique* 50. 79–95.

Abeillé, Anne, Olivier Bonami, Danièle Godard & Jesse Tseng. 2006. The syntax of French à and de: an HPSG analysis. In Patrick Saint-Dizier (ed.), *Syntax and semantics of prepositions*, 147–162. Springer.

Anttila, Arto, Matthew Adams & Michael Speriosu. 2010. The role of prosody in the English dative alternation. *Language and Cognitive Processes* 25(7-9). 946–981.

Artstein, Ron. 2005. Coordination of parts of words. *Lingua* 115(4). (Coordination: Syntax, Semantics and Pragmatics). 359–393.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2014. *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-6*. http://CRAN.R-project.org/package=lme4.

Boersma, Paul. 1998. *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.

Booij, Geert E. 1985. Coordination reduction in complex words: a case for prosodic phonology. In Harry Van der Hulst & Norval Smith (eds.), *Advances in Nonlinear Phonology*, 143–160. Dordrecht: Foris.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen. 2007. Predicting the dative alternation. In G Boume, I Kraemer & J Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Chaves, Rui P. 2008. Linearization-based word-part ellipsis. *Linguistics and Philosophy* 31(3). 261–307.

Cohn, Abigail. 2006. Is there gradient phonology? In Gisbert Fanselow, Caroline Féry, R Vogel & Matthias Schlesewsky (eds.), *Gradience in grammar: generative perspectives*, 25–44. Oxford: Oxford University Press.

---

[7] Abeillé (2006) also claims that in French, prefixes can't take wide scope at all: **Il ne faut pas surévaluer et ~~sur~~noter* 'One should not overevaluate and ~~over~~-rate'. That is, the utterance is legal, but cannot have the meaning of *over-rate*, only of *rate*. In English, however, word-initial deletion requires a strong supporting context to be interpreted as such, and perhaps the same is true in French. In isolation, *step-sons and daughters* would probably not be interpreted as involving ~~step~~-*daughters*, but in the context *When Maria married Captain von Trapp, she acquired several step-sons and daughters*, the ~~step~~-*daughters* reading is salient if the listener is familiar with the movie *The Sound of Music*, in which then-childless Maria marries a widowed man with several sons and daughters.

Embick, David & Rolf Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32(4). 555–595.

Green, Thomas & Michael Kenstowicz. 1995. The Lapse Constraint. In Leslie Gabriele (ed.), *Proceedings of the Sixth Annual Meeting of the Formal Linguistic Society of Mid-America*, vol. I, 1–14. Bloomington IN: Indiana University Linguistics Club.

Grevisse, Maurice. 1980. *Le bon usage*. Gembloux: Duculot.

Hadfield, Jarrod D. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* 33(2). 1–22.

Hayes, Bruce & Zsuzsa Cziráky Londe. 2006. Stochastic Phonological Knowledge: The Case of Hungarian Vowel Harmony. *Phonology* 23(01). 59–104.

Lin, Yuri, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman & Slav Petrov. 2012. Syntactic Annotations for the Google Books Ngram Corpus. *Proceedings of the ACL 2012 System Demonstrations*, 169–174. (ACL '12). Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=2390470.2390499.

Löfstedt, Ingvar. 2010. Phonetic Effects in Swedish Phonology: Allomorphy and Paradigms. UCLA Ph.D. Dissertation.

Matos, Rafael, Ludovic Ferrand, Christophe Pallier & Boris New. 2001. Une base de données lexicales du français contemporain sur internet : LEXIQUE. *L'année psychologique* 101(3). 447–462.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, Dale Hoiberg, et al. 2011. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331(6014). 176–182.

Miller, Philip H. 1992. *Clitics and Constituents in Phrase Structure Grammar*. New York: Garland.

New, Boris, Christophe Pallier, Marc Brysbaert & Ludovic Ferrand. 2004. Lexique 2: a new French lexical database. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc* 36(3). 516–524.

Pierrehumbert, Janet. 1994. Knowledge of variation. *Papers from the Parasession on Variation, 30th meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.

Prince, Alan & Paul Smolensky. 1993. *Optimality Theory*. Blackwell.

Pullum, Geoffrey & Arnold Zwicky. 1988. The syntax-phonology interface. In F. Newmeyer (ed.), *Linguistics: The Cambridge Survey*, vol. 1, 255–280. Cambridge: Cambridge University Press.

Raffelsiefen, Renate. 1999. Phonological constraints on English word formation. In Geert E Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1998*, 225–287. (Yearbook of Morphology 8). Springer.

R Core Team. 2014. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. www.R-project.org.

Shih, Stephanie. 2014. Towards optimal rhythm. Stanford University PhD dissertation.

Stodden, Victoria, Friedrich Leisch & Roger D Peng (eds.). 2014. *Implementing Reproducible Research*. (Chapman & Hall/CRC The R Series). Chapman and Hall/CRC.

Tranel, Bernard. 1996. French liaison and elision revisited: A unified account within Optimality Theory. In Claudia Parodi, Carlos Quicoli, Mario Saltarelli, Maria Luisa Zubizarreta, Claudia Parodi, Carlos Quicoli, Mario Saltarelli & Maria Luisa Zubizarreta (eds.), *Aspects of Romance Linguistics*, 433–455. Washington, DC: Georgetown University Press.

Tseng, Jesse. 2003. Phrasal affixes and French morphosyntax. In G Jäger (ed.), *FGVienna: Proceedings of Formal Grammar 2003*, 177–188. Technische Universität Wien.

Wagner, Michael. 2011. Production planning constraints on allomorphy. *Journal of the Canadian Acoustical Association* 39(3). 160–161.

Xie, Yihui. 2014a. *knitr: a general-purpose package for dynamic report generation*.

Xie, Yihui. 2014b. knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch & Roger D Peng (eds.), *Implementing Reproducible Computational Research*. Chapman and Hall/CRC.

Yip, Moira. 1995. Repetition and its Avoidance: The Case in Javanese. In Keiichiro Suzuki & Dirk Elzinga (eds.), *Proceedings of South Western Optimality Theory Workshop 1995*, 238–262. University of Arizona, Department of Linguistics.