David Lorenz* and David Tizón-Couto

Chunking or predicting – frequency information and reduction in the perception of multi-word sequences

<https://doi.org/10.1515/cog-2017-0138>

Received 15 November 2017; revised 10 April 2019; accepted 13 April 2019

Abstract: Frequently used linguistic structures become entrenched in memory; this is often assumed to make their consecutive parts more predictable, as well as fuse them into a single unit (chunking). High frequency moreover leads to a propensity for phonetic reduction. We present a word recognition experiment which tests how frequency information (string frequency, transitional probability) interacts with reduction in speech perception. Detection of the element *to* is tested in *V-to-V_{inf}* sequences in English (e.g., *need to V_{inf}*), where *to* can undergo reduction (“needa”). Results show that reduction impedes recognition, but this can be mitigated by the predictability of the item. Recognition generally benefits from surface frequency, while a modest chunking effect is found in delayed responses to reduced forms of high-frequency items. Transitional probability shows a facilitating effect on reduced but not on full forms. Reduced forms also pose more difficulty when the phonological context obscures the onset of *to*. We conclude that listeners draw on frequency information in a predictive manner to cope with reduction. High-frequency structures are not inevitably perceived as chunks, but depend on cues in the phonetic form – reduction leads to perceptual prominence of the whole over the parts and thus promotes a holistic access.

Keywords: speech perception, phonetic reduction, chunking, frequency information, entrenchment

1 Introduction

In spoken language use, phonetic reduction is ubiquitous, full canonical articulation of words is the exception. Among the factors that bring about reduction

*Corresponding author: David Lorenz, Institut für Anglistik/Amerikanistik, Universität Rostock, Rostock, Germany, E-mail: david.lorenz2@uni-rostock.de

David Tizón-Couto, Department of English, French and German, Universidade de Vigo, Vigo, Spain, E-mail: davidtizon@uvigo.es

are frequency-based probabilities; generally, words are likely to be reduced when they occur in frequent phrases and are predictable from the linguistic context (cf. Gregory et al. 1999; Jurafsky et al. 2001; Bell et al. 2003, Bell et al. 2009; Gahl and Garnsey 2004; Warner and Tucker 2011; Gradoville 2017). Reduction can also occur across word boundaries. For example, certain multi-word sequences undergo phonological reduction and contraction to a single word (e.g., *want to* > *wanna*). In usage-based approaches, this is seen as a consequence of entrenchment, resulting from frequency of occurrence. There are, broadly, two different ways of seeing this. Firstly, entrenchment may be a form of “procedure strengthening” (cf. Hartsuiker and Moors 2018), that is, an automatization of a sequence of separate items that is frequently encountered and repeated, and is therefore predictable. This can be explained in terms of probabilistic learning (speakers/hearers have a tacit knowledge of sequential probabilities) and in terms of syntagmatic associations (speakers/hearers associate the two items due to their frequent co-occurrence). The second, perhaps more prominent view is that frequent phrases and sequences undergo “chunking” (Bybee 2002, Bybee 2006; Ellis 2002; Diessel 2007; Ellis et al. 2009). Thus, frequent multi-word sequences will be stored in the mind as a single unit and their component parts are backgrounded (cf. Langacker 2000: 278). Sequences of this kind (e.g., *want to*) have a propensity for reduction due to neuromotor routines and automatized articulation (Bybee 2006). In turn, reduced forms can be further reinforced by frequency and result in contractions (such as *wanna*). Thus, the mental representation of such multi-word units may also come with pronunciation variants, in the same way as reduced variants of single words are stored in memory (cf. Patterson et al. 2003; Bürki et al. 2011; Seyfarth 2014). Presumably, reduced forms are more or less strongly represented in the language user’s mind on a gradient cline that ranges from outcomes of on-line articulatory reduction (with no prior representation) to fixed variants that are stored in memory (cf. Connine and Pinnow 2006; Lorenz 2013; Tizón-Couto and Lorenz 2018).

Most of the existing evidence of chunking and the reducing effect of frequency has dealt with language production only, which raises the question of how they might affect speech perception. Regarding reduction in individual words, there is some evidence that full canonical forms generally serve the listener best in auditory lexical decision tasks (Ernestus and Baayen 2007; Ranbom and Connine 2007; Pitt 2009; Tucker 2011; Pitt et al. 2011), and that recognition deteriorates with increasing reduction (Ernestus et al. 2002). There is also conflicting evidence that it is a word’s most frequent variant form – rather than the canonical form – that generates greater lexical activation in experimental tasks (Connine 2004; Connine et al. 2008; Bürki and Frauenfelder 2012;

Racine et al. 2014; Bürki et al. 2018). Taken together, these studies show that the connection between item frequency, variant frequency, reduction and recognition is not straightforward.

Listeners also apply a knowledge of word sequences, such that, for example, high-frequency phrases are recognized faster (Arnon and Snider 2010). Again, such phrases may be easy to process because they form a known and expectable combination of items, or because they are treated as single items due to chunking. Evidence for chunking in speech perception has been presented by Sosa and MacFarlane (2002). In a word recognition experiment, they find delayed responses to elements of highly frequent sequences (e.g., *of* in *sort of*). They conclude that “[a]ccessing *of* as a constituent of *kind of* [...] might entail a process of morphological decomposition, or it might require the use of explicit language knowledge” (Sosa and MacFarlane 2002: 234). Thus, listeners perceive the bigram as a single unit before identifying its elements. Their design did not, however, consider that these sequences have a propensity for reduction (e.g., “kinda” [ˈkɑndə]) and that this might have an effect on word recognition. In a similar study investigating the recognition of *up* in collocations of differing frequencies (e.g., *sign up*, *run up*), Kapatsinski and Radicke (2009) report a U-shaped effect: word recognition is delayed in sequences of both very high and very low frequency. They suggest an interplay of procedure retrieval and stored representation of a “chunk”. Frequent co-occurrence increases the predictability of a word, hence facilitates its recognition; however, this facilitating effect seems to be offset in collocations of very high frequency due to chunking and low perceptual salience. This corresponds well with a notion of chunking as a gradual process by which the whole takes precedence over its parts, that is, the single-unit representation is more readily retrieved but does not fully block activation of the individual parts and their composition (cf. Blumenthal-Dramé 2018: 130, 138). Overall, however, the role of chunking vis-à-vis procedure strengthening in speech perception is less than well-understood (see Divjak and Caldwell-Harris 2015 for a more general discussion). Among the open questions are: To what extent do chunking and procedure strengthening follow from high surface frequency, or from other frequency measures (such as conditional probabilities)? What are the roles of chunking and procedure strengthening in recognizing phonetically reduced forms?

The present study addresses these questions; it builds on the experiments by Sosa and MacFarlane (2002), and Kapatsinski and Radicke (2009), with two important additions. Firstly, it considers the effect of reduction in the speech input; secondly, the effect of frequency information is tested not only in terms of surface frequency but also conditional probability. We conducted a word recognition experiment using the function word *to* in V-*to*-V_{inf} constructions in

American English (e.g., *have to* V_{inf} , *prefer to* V_{inf}). This construction comes with a range of main verbs of different frequencies and probabilities, and has a potential for chunking and reduction of *to* (/t/-lenition and vowel reduction).

We may expect that phonetic reduction of the target word generally impedes its recognition. The crucial question is how frequency information (string frequency and conditional probability) and reduction ([tʊ] > [rə]) interact. String frequency could either aid or hinder the detection of reduced items. On the one hand, given that high frequency strengthens the syntagmatic associations within the sequence, listeners may have an active knowledge of the high probability of *to* based on frequency. The item, and perhaps also its appearance as a reduced form, may then be more expected and reduction more easily compensated (cf. Jurafsky et al. 2001; Arnon and Cohen Priva 2013; Van de Ven and Ernestus 2018); in this case reduction would affect recognition times less as frequency increases. On the other hand, listeners may have a chunked variant available for highly frequent bigrams; in that case a reduced form would lead them to access this holistic representation and considerably delay the recognition of *to*. These two scenarios need not be mutually exclusive, if chunking only sets in at very high frequencies.

Conditional probability, i.e., the likelihood of a word given its context (e.g., the previous word), is a frequency-based cue which has not received as much attention as surface frequency, especially with respect to chunking. In speech production, conditional probability has been claimed to affect phonetic realization: it has an effect on both the duration and vowel quality (full vs. reduced) of function words such as *the*, *of*, or *to* (Jurafsky et al. 2001; Bell et al. 2003: 1018, Bell et al. 2009: 102–103). However, these findings imply no claim that conditional probability might lead to chunking. In this sense, the role of conditional probability in fusing the bigram into a single unit remains an open question, especially vis-à-vis the oft-reported chunking effect of string frequency.

On the perception side, high conditional probability facilitates word recognition. This has been shown in reading experiments (McDonald and Shillcock 2003; Tremblay and Tucker 2011), as well as for auditory speech perception (e.g., Simpson et al. 1989; Frank and Willems 2017). In these accounts, conditional probability serves as a contextual cue to predicting upcoming words.¹ Listeners

¹ While the present study considers word-based transitional probabilities, other contextual cues come from lexical-semantic prediction of candidate words (e.g., Altmann and Kamide 1999; Van Berkum et al. 2005; see Simpson et al. 1989; Frank and Willems 2017 for differentiation of semantic and sequential predictions), and from transitional probabilities of syllables for identifying word boundaries (Fernandes et al. 2007; Asheimer and Sanders 2011; Franco and Destrebecqz 2012).

appear to maintain such contextual information by default (Bushong and Jaeger 2017), but its tangible effect on word recognition may be limited. Mattys et al. (2012) suggest that listeners especially rely on contextual cues when the input signal is less clear, such as with noise or phonetic reduction in casual speech. Invoking cues and prediction assumes that listeners apply probabilities rather directly, which clearly corresponds to the cognitive mechanisms of syntagmatic associations and procedure strengthening.

The facilitating effect of conditional probability could interact with reduction in two possible ways. On the one hand, listeners may generally draw on the information provided by conditional probability in order to identify *to*, regardless of form (full or reduced); in this case, recognition would be equally affected by probability for each variant. On the other hand, the information from conditional probability may become more relevant when the input signal is less informative (in this case, phonetically reduced). Thus, a high conditional probability would particularly facilitate recognition of a reduced item, but not affect full forms as much.

The word recognition experiment reported here examines the effect of frequency-based cues in the processing of multi-word sequences in speech. This is measured by recognition accuracy and response times to the element *to* in *V-to-V_{inf}* constructions, when the element is either fully articulated or phonetically reduced. The analysis considers frequency information in terms of surface frequency and transitional probability. Several control variables are included which show that recognition is also affected by phonological properties of the items.

2 Experiment design

2.1 Participants and stimuli

The data come from 38 native speakers of American English (20 female), who were offered a compensation for their participation. The stimuli consist of 126 recorded sentences in American English, all produced by the same speaker. These comprise 42 target items containing a *V-to-V_{inf}* construction, each with a different verb before *to* (1); 42 control items containing *to* in a different construction (2); and 42 distractors which do not contain *to* at all (3). The target items were designed to ensure that the subject was always a pronoun and the target word *to* occurred in the middle of the sentence. Control items contained the same verbs as the target items, but not in a *V-to-V_{inf}* sequence, so that

participants could not simply assume that a verb would be followed by *to* whenever plausible.

- (1) When the penguins are around, we pretend to like the way they dress.
- (2) When the monkeys come over to dance, I pretend I'm asleep.
- (3) I can't believe the crocodiles are selling leather handbags.

Each target item was recorded in two conditions: “full” and “reduced”, referring to the pronunciation of *to*. Full pronunciations consisted of a full [t] and a short [ʊ]; reduced forms had a flap [ɾ] and a schwa [ə] (e.g., *pretend to* as “pretenda”). The recordings were checked to make sure that *to* could be distinguished as full or lenited on the basis of both auditory impression and acoustic analysis (waveform).² Speaking rate in the test sentences was held constant at around 5 syllables per second.³

Each participant responded to 126 items in total, 42 each of target, control and distractor items. The target items all contained different instances of *V-to-V_{inf}*, 21 of which were presented in the “reduced” condition and 21 as full forms. Participants were assigned to one of two groups, so that group A would hear full forms of those items that were reduced for group B, and vice versa (see Appendix for a list of the items). The control items were the same for each group and all presented with a full *to*.

2.2 Design and procedure

Participants were asked to respond to the presence or absence of *to* as accurately and quickly as possible by pressing a key on the keyboard. They were told that “*to* is a very common and versatile word in English. It can serve many functions, and it is not always pronounced the same way”. The experiment was preceded by a brief practice phase. Participants answered a simple comprehension question on 10 of the control items during the experiment to confirm their continuing attention to the stimuli. The order of the stimuli was randomized.

The software *OpenSesame* (version 3.0.7 for Mac, Mathôt et al. 2012) was used to carry out the experiment and record response times. By using the

² We used *Praat* (version 6.0.16, Boersma and Weenink 2016) for this and all other acoustic measurements (durations, /t/-onset).

³ It may be argued that reduction would appear more natural in rapid speech, and that speaking rate can affect listeners' expectations (cf. Brown et al. 2012). By keeping speaking rate variance to a minimum, we control for such expectations. At the same time, if reduction follows from frequency-based chunking, it should be more likely in high-frequency items regardless of speaking rate (cf. Tizón-Couto and Lorenz 2018).

“xpyriment” backend, the experiment runs on the Python library *Expyriment* (Krause and Lindemann 2014), which is recommended due to its timing precision. The “prepare-run strategy” implemented in *OpenSesame* (i.e., preparing a stimulus before running it) further prevents timing errors due to other hardware operations. Since the experiment was run on different computers in different locations, the temporal jitter was tested on each machine by 30 random trials; the deviations did not exceed ± 10 milliseconds.

2.3 Variables and analysis

2.3.1 Dependent variables: response time and accuracy

Response times were measured from the onset of *to*. The reference point for the onset of *to* is the release burst of the plosive, since this point can be more reliably established than the preceding closure: independent measurements by the two authors only differed by 1–5 milliseconds.

After removing distractor items, we had 3,192 data points for control and target items. Data points with wrong responses (i.e., where *to* was not identified) or implausible response times were considered inaccurate. We initially left a window of 100 – 3,000 ms for this first a priori data selection. Outlier values were then identified through a mild by-subject screening of the data (Baayen and Milin 2010: 16), so that values with a by-subject *z*-score of >2.5 standard deviations were also labeled inaccurate. After this procedure, the data comprise 1,367 correct responses on target items (1,596 total; accuracy 86%). The accuracy rate for reduced stimuli (79%; 630/798) is lower than for the full items (92%; 737/798).

Our main analysis concerns response times on correctly identified items, as these typically reflect processing difficulty (Section 3.2). However, given the difference in accuracy rates, we also consider accuracy as a dependent variable in a separate model (Section 3.1).

2.3.2 Independent variables

The independent variables we considered are the following⁴:

⁴ The test variables “frequency” and “transitional probability” show a slight positive correlation (Spearman’s $\rho = 0.375$), but there is no indication of any problematic collinearity. Omitting one from the model does not strongly change the slopes and smoothing of the other

Condition

The experimental conditions are a full or reduced realization of *to* ([tʊ] vs. [rə]) in the *V-to-V_{inf}* sequence in target items. We are interested in seeing how reduction influences the effects of other variables; therefore, condition is used as a moderator variable in the analysis.

Frequency

The *V-to-V_{inf}* sequences in the target items are of varying frequencies, as derived from the surface frequencies per 1 million words in the spoken section of the Corpus of Contemporary American English (COCA, Davies 2008). Surface frequencies are the frequency of the given verb form with a *to*-infinitive complement. For each verb, the inflected form with the highest surface frequency was chosen (for example, the progressive in *trying to V_{inf}*, past tense in *began to V_{inf}*). These forms are most likely to show frequency effects without interference from more common forms of the same verb (e.g., the perception of *try to* being confounded by an expectation of *trying to*).

We must expect frequency to have a non-linear effect, and a different effect on full and reduced items. Previous studies have dealt with this issue by grouping frequencies into bins (Sosa and MacFarlane 2002; Kapatsinski and Radicke 2009). In our analysis (Section 3), we maintain frequency as a continuous variable and use smooth terms that do not pre-suppose any particular shape of the data. Examples (4)–(7) illustrate the frequency range from very high (4) to very low (7).

- (4) If the camel is sick, we have to give him his medicine.
- (5) The way the penguins dress they seem to have a lot of money.
- (6) When the giraffe is here we intend to give him a room upstairs.
- (7) Surely, you would never deign to play chess with my camel.

Transitional probability of *V > to*

Transitional probabilities for each *V > to* sequence were also established on the basis of the spoken section of COCA. Transitional probability (TP) measures the likelihood of *to* occurring after a particular verb; it is calculated by dividing the frequency of the bigram (*V to*) by the frequency of the first element (*V*).

By way of example, the sequence *deign to* has one of the highest transitional probabilities (despite its low frequency), while *have to* ranks low (despite its high frequency). As with surface frequency, the effect of transitional probability may be non-linear and interact with condition. It will therefore be analyzed with smooth terms in the same way as the frequencies.

(cf. Zuur et al. 2009: 66); variance inflation factors for the independent variables are all below 2 (checked using `corvif()`, Zuur et al. 2009: 255).

Frequency of to-V_{inf} and backward transitional probability

Phonetic reduction can also be conditioned by the item's co-occurrence frequency with the following word (cf. Bell et al. 2009; Barth and Kapatsinski 2017; Gradoville 2017). In perception, a word that often occurs after *to* might aid the hearer to more quickly realize that they just heard the element *to*. We check for this by considering both the surface frequencies of *to-V_{inf}* sequences and the probability of *to* given the following word (backward transitional probability). These measures were established from the spoken section of COCA. Since the experimental items were designed to avoid surprises at the *V_{inf}* position, the set of verbs following *to* is limited to inconspicuous ones such as *play*, *build*, or *give*.

Verb duration, syllable count and verb form

The duration of the verb *preceding to* was calculated by subtracting the timing for the onset of *to* from the timing for the onset of the verb. Verb durations range from 182 to 590 ms (mean = 350 ms). Since many verbs are monosyllabic, an additional factor considers whether the verb has one syllable or more.

The inflectional forms of the Vs in the target items were chosen by selecting for each verb the form with the highest surface frequency in the corpus. To control for the potential influence of inflection, the items were coded for present, past (*-ed*, *chose*, *began*) and progressive (*-ing*). Out of the 42 verbs, 26 were in the base form (present), 12 in past tense, and 4 in progressive.

Merged plosive cluster

Initial Vs might end in an alveolar stop (/d/ or /t/), e.g., in *need to* or *forgot to* (8)–(9). In such cases, this sound merges with the initial /t/-sound in *to*, regardless of its shape (full or reduced). The separation between the verb and *to* is less clear in these items. This plosive merger is present in 19 verbs in our set.

(8) Careful now, we need to watch out for monkeys around here.

(9) Yesterday I forgot to give the camels their food.

Preceding sound

In speech production, lenited /t/-sounds are typically disfavored after fricatives (cf. Lorenz and Tizón-Couto 2017). Some of our experimental items include lenition in this context. The sound segment preceding /t/ (i.e., the last sound in the V) was coded for two levels: fricative (e.g., *have to*) or vowel/nasal (e.g., *agree to*, *happen to*). A fricative in this position is found in 10 of the verbs.

Gender and age

Participants were asked to provide their details as regards these two variables in one of the screens introducing the experiment. Age was taken up by year of birth

and ranges from 1952 to 1997 (mean = 1985); it was centered and scaled by standard deviation for the analysis.

Control before target and item count

The control items contained the same verbs as the target items, but with a different complement (see example (2) above). Since the stimuli were presented in random order, this variable checks whether a participant heard the corresponding control item before the target item.

To control for learning or fatigue effects, the item count during the experiment was considered, both as a control variable and as a random effect varying for each participant.

3 Results

In order to assess how accuracy and response times are conditioned, we fitted a mixed-effects generalized additive model (GAM; cf. Wood 2006; Zuur et al. 2009: 35–70; Wright and London 2009: 112–137) for each.⁵ “Accuracy” is a binary variable (correct vs incorrect response) and thus requires a binomial model (Section 3.1); response times have been logarithmically transformed and modeled as a continuous variable (Section 3.2). Both models include smooth terms (cubic regression splines) for the test variables “frequency” (log-transformed) and “transitional probability” (TP). The control variables listed above are included as parametric factors. “Condition” (full vs reduced) serves as a moderator variable, that is, the interaction of “condition” with every other independent variable is considered. Since the effects of “frequency” and “transitional probability” may be non-linear, the smooth terms ensure that these can be modeled without prior assumptions of their shape. Due to the interaction with “condition”, the model fits a separate curve for each condition. Thus, it shows how the effects of frequency and transitional probability differ between full and reduced input items.

When modeling mixed effects, smooth terms, interactions and control variables, the question of variable selection is a complex one for which no fool-proof standard procedure exists (see Barr et al. 2013; Bates et al. 2015; Baayen et al. 2017 for discussion on random effects and parsimony). The modeling procedure applied here is based on backward stepwise variable selection for both random and fixed effects. We proceeded as follows.

⁵ All data analysis was done in R (R Core Team 2017). We used the function `gam()` from the R package *mgcv* (Wood 2006, 2011) to create the statistical models; and `compareML()` from *itsadug* (Van Rij et al. 2017) for model comparisons. Our data sets and R code for the statistical analyses and graphs are available at the TROLLing data repository (<https://doi.org/10.18710/7TSABU>).

First, the random effects structure was determined by creating a model with only the test variables (“frequency” and “TP”, each interacting with “condition”) and a maximal random effects structure. This structure controls for individual differences between participants by random intercepts for “subject” and by-subject random slopes on “item count”, “condition”, “frequency” and “TP”; and for idiosyncrasies of particular test items with random intercepts for “verb” and by-verb random slopes on “condition”. We successively eliminated those random factors that do not make a significant contribution to the model based on *AIC* and *ANOVA* comparison.

In the next step, we used the resulting random effects structure in a model that includes the test variables and all control variables. Random effects address the “human factor” and experimental noise in the data. By setting up the random effects structure first, the other variables will only show the effect they have beyond the variance that is controlled for by the random effects, that is, we reduce the danger of measuring noise as a fixed effect. Control variables (including interactions with “condition”) were successively eliminated until the exclusion of any further term significantly weakened the model (based again on *AIC* and *ANOVA* comparison).

The resulting models for accuracy and response times thus comprise different variables, but are comparable in their interpretation. We first present the results on accuracy (Section 3.1) and then focus on response times (3.2).

3.1 Accuracy

The overall accuracy rate on the 1,596 target items is 85.7%. Full renderings of *V-to-V_{inf}* were recognized more consistently (accuracy 92.4%) than the reduced variants (78.9%). On control items, the accuracy rate is 88.6%. Thus, as expected, full target items do not differ significantly from control items, while reduced forms produce more errors.

To see whether this difference is influenced by frequency, conditional probability or other variables, a binomial generalized additive model (GAM) was fitted following the procedure set out above. The resulting model includes smooth terms for the test variables frequency (“logfreq”) and transitional probability (“TP”); for “TP”, the interaction with “condition” turned out to be irrelevant and is excluded from the model. Fixed effects are included for three control variables: “plosive cluster”, i.e., a final /t/ or /d/ in the verb preceding *to*; “control before target”, i.e., whether the verb had already been heard in a control item; and the item count during the experiment (included as a linear effect because a smooth term provided no clear advantage to the model). Random effects are present as intercept adjustments for “subject” and “verb” (the test item). The model is specified as

$$\begin{aligned} \text{correct} \sim & \text{s}(\text{logfreq}, \text{bs} = \text{"cr"}, \text{by} = \text{condition}) + \text{s}(\text{TP}, \text{bs} = \text{"cr"}) + \\ & \text{condition} * \text{plosive_cluster} + \text{condition} * \text{control_before_target} + \\ & \text{condition} * \text{item_count} + \\ & \text{s}(\text{subject.fac}, \text{bs} = \text{"re"}) + \text{s}(\text{verb}, \text{bs} = \text{"re"}) \end{aligned}$$

The model provides a good fit to the data ($C = 0.836$, $UBRE = -0.301$)⁶; it is presented in Table 1 and visualized in Figure 1.⁷ It should be noted that since

Table 1: GAM model summary (Accuracy).

Parametric coefficients:					
	Estimate	Std. Error	t	p	significance
(Intercept)	-2.151	0.355	-6.064	< 0.001	***
condition = "reduced" (reference level = "full")	1.188	0.362	3.278	0.001	**
plosive cluster = "yes" (reference level = "no")	-0.940	0.398	-2.365	0.018	*
control before target = "yes" (reference level = "no")	-1.109	0.368	-3.009	0.003	**
item count (continuous)	0.002	0.005	0.426	0.670	
Interaction					
condition : plosive cluster "reduced" : "yes"	1.447	0.406	3.562	< 0.001	***
Interaction					
condition : control before target "reduced" : "yes"	1.178	0.446	2.639	0.008	**
Interaction					
condition : item count "reduced" : "continuous"	-0.016	0.006	-2.765	0.006	**
Smooth terms for test variables:					
	edf	Ref. df	F	p	significance
frequency (log) : condition = full	1.000	1.001	0.117	0.733	
frequency (log) : condition = reduced	6.279	6.917	13.229	0.057	.
TP	1.002	1.002	2.456	0.117	
[Random effects (smooth terms)]					
subject	26.068	37.000	81.072	< 0.001	***
verb	23.694	38.000	63.504	< 0.001	***

6 As with logistic regression models, C is a concordance index, where $C \geq 0.8$ indicates a good fit. $UBRE$ is the “unbiased risk estimator” that is minimized in the smooth term estimation (Wood 2006: 175–179); similar to AIC , it is only meaningful in model comparison.

7 The upper row of Figure 1 presents the visual `gam()` output; the lower panels were created with `ggplot2` (Wickham 2016) and functions from `Rmisc` (Hope 2013).

Figure 1: Effects on accuracy. Y-axes for smooth terms (upper row) refer to estimated effects, which are centered and scaled by the GAM. Y-axes for parametric terms (lower row) refer to model predictions of error rate. The upper panels can only be read individually (i.e., 0 does not refer to an overall mean). The effective degrees of freedom (*edf*) provide a measure of the (non-)linearity of the smoothing curve – an *edf* of 1 stands for a linear effect, while a higher *edf* means more non-linearity (i.e., a wigglier curve; cf. Zuur et al. 2009: 53).

accuracy is fairly high overall, the results it provides are rather coarse-grained.

According to the model, frequency and transitional probability have little impact on the occurrence of errors. There are slightly fewer recognition errors with high TP, but this trend is not significant ($p = 0.117$). Frequency is marginally significant only on reduced items ($p = 0.057$). While it should be viewed with caution, the smooth curve (Figure 1, top middle) suggests that items of mid-high frequency are recognized with the fewest errors. Moreover, the presence of a plosive merger increases the error rate on reduced items (Figure 1, bottom left).

These first results already hint at a possible facilitating role of probability; the difficulty of recognizing reduced forms is aggravated when reduction affects a morpheme boundary (as with plosive clusters), and may be mitigated at mid-high levels of frequency and high TP. Measuring accuracy can only provide a rough measure of these effects, and the smooth terms only yield statistical trends. It will be seen that similar effects hold more clearly with the response times on correctly recognized items. In addition, participants made fewer errors when they had heard the same verb in a different construction (i.e., a control

item) before the $V\text{-}to\text{-}V_{\text{inf}}$ item; error rates on reduced forms in particular decrease in the course of the experiment (“item count”). This suggests that there are learning effects as participants get attuned to the voice and content of the stimuli. While the learning effect applies to accuracy, we will see that it does not translate into response times.

3.2 Response times

The overall mean response time was 675 ms (median: 580 ms). As the bean plots in Figure 2 show, responses to full renderings of $V\text{-}to\text{-}V_{\text{inf}}$ are just slightly faster than to control items ($t = -1.997, p = 0.046$), whereas a reduced *to* elicits considerably longer response times ($t = 8.924, p < 0.001$).⁸ Thus, as expected, reduction of the target word generally delays its recognition. The crucial question is to what extent this delay may be mitigated by frequency, conditional probability or other, phonological cues.

Figure 2: Response times to control, full and reduced items.

We fitted the log-transformed response times of target items ($n = 1,367$) to a mixed-effects generalized additive model following the procedure set out above. With “condition” (full vs reduced) as a moderator variable, the model tests how the effect of reduction is affected by the other variables. The model includes smooth terms for the test variables “frequency” (log-transformed) and “transitional probability”,⁹

⁸ We used the R package *beanplot* (Kampstra 2008) to create the graph; the shape of the beans represents the data distribution, horizontal lines refer to group means. The t - and p -values reported are from a simple linear model of log response times by condition, with “control” as reference level. A likely explanation for the slower responses to control items (compared to full target items) is that the particle *to* is not consistent in function and position in control sentences.

⁹ The default basis dimension of $k = 10$ was used for the smooth terms, so that the model finds the best-fitting smooth with maximally 10 knots and maximally 9 *edf*. As the resulting smooths have *edf* far below this maximum, and lowering k to enforce (near-)linear effects leads to less accurate models (by *GCV* comparison), there is no indication that the k settings need to be adjusted (see Wood 2006: 159).

and parametric terms for the control variables “verb duration”, “plosive cluster”, “age” and “gender”. Random intercepts are included for “subject” and “verb” as well as random smooths for item count (by subject) and condition (by verb). The final model is thus defined as:

$$\begin{aligned} \text{logrt} \sim & \text{s}(\text{logfreq}, \text{bs} = \text{"cr"}, \text{by} = \text{condition}) + \text{s}(\text{TP}, \text{bs} = \text{"cr"}, \text{by} = \text{condition}) + \\ & \text{condition} + \text{log}(\text{verb_duration}) + \text{condition} * \text{plosive_cluster} + \text{condition} \\ & * \text{gender_response} + \text{condition} * \text{age.scaled} + \text{s}(\text{subject.fac}, \\ & \text{bs} = \text{"re"}) + \text{s}(\text{verb}, \text{bs} = \text{"re"}) + \text{s}(\text{item_count}, \text{subject.fac}, \text{bs} = \text{"re"}) \\ & + \text{s}(\text{condition}, \text{verb}, \text{bs} = \text{"re"}) \end{aligned}$$

Overly influential outliers were eliminated post-hoc through the model residuals, removing data points with a residual value beyond 2.5 standard deviations (33 items; cf. Baayen and Milin 2010: 17–18). After this trimming, the final model stands up to scrutiny (near-normal distribution and constant variance of residuals, no extreme outliers) and reaches $R^2 = 0.612$.¹⁰ It is presented in Table 2.

Table 2: GAM model summary (Response Times).

Parametric coefficients:					
	Estimate	Std. Error	t	p	significance
(Intercept)	8.176	0.561	14.579	<0.001	***
condition = “reduced” (reference level = “full”)	1.113	0.041	2.815	0.049	**
verb duration (log) (continuous)	−0.314	0.091	−3.481	<0.001	***
plosive cluster = “yes” (reference level = “no”)	−0.025	0.049	−0.504	0.614	
gender = “male” (reference level = “female”)	0.002	0.268	0.01	0.992	
age (scaled) (continuous)	−0.111	0.135	−0.818	0.414	
Interaction					
condition : plosive cluster “reduced” : “yes”	0.209	0.056	3.746	<0.001	***

(continued)

¹⁰ Adjusted $R^2 = 0.575$; $GCV = 0.096$. The GCV (generalized cross-validation score) is minimized in the smooth term estimation (Wood 2006: 175–179); similar to AIC , it is only meaningful in model comparison.

Table 2: (continued)

Parametric coefficients:					
	Estimate	Std. Error	t	p	significance
<i>Interaction</i>					
condition : gender	−0.064	0.034	1.89	0.061	.
“reduced” : “male”					
<i>Interaction</i>					
condition : age (scaled)	−0.042	0.017	−2.456	0.015	*
“reduced” : “age (scaled)”					
Smooth terms for test variables:					
	edf	Ref. df	F	p	significance
frequency (log) : condition = full	1.000	1.000	12.576	< 0.001	***
frequency (log) : condition = reduced	3.882	4.114	2.274	0.055	.
TP : condition = full	1.000	1.000	0.162	0.688	
TP : condition = reduced	1.000	1.000	11.364	< 0.001	***
[Random effects (smooth terms)]					
subject	34.676	35.000	103.286	< 0.001	***
verb	22.416	39.000	9.088	< 0.001	***
item count: subject	15.429	38.000	245.635	< 0.001	***
condition: verb	29.091	76.000	2.145	< 0.001	***

3.2.1 Test variables

In the model in Table 2, “condition” shows the expected effect of reduced forms leading to longer response times. The smooths in Figure 3 show how the effects of frequency (“logfreq”) and “TP” also vary between conditions.

For frequency, the smooth terms in the model show that there is a strong (and linear) effect on the full forms, while the effect on reduced forms barely reaches significance. As the curves in Figure 3 (upper panels) show, low-frequency items produce longer response latencies in both conditions; recognition of full forms continuously profits from higher frequency, producing a linear effect ($edf=1$, $p < 0.001$). Response times to reduced forms rather follow a shallow U-shaped curve, which verges on statistical significance ($edf=3.82$, $p = 0.055$). Higher frequencies lead to faster responses up to a point at around $\logfreq=3$, but this trend does not continue at very high frequencies.

Figure 3: Effects of surface frequency and transitional probability on response times.

For transitional probability, there is a very clear interaction with “condition” (Table 2). An increase in TP does not make a significant difference to recognition of *to* in its full form, but it has a clear impact on the reduced items ($p < 0.001$). The effect is linear ($edf = 1$), such that recognition of reduced forms is continuously faster with increasing TP.

The differences in recognition latencies between full and reduced forms at different frequencies and TPs are shown in Figure 4. Here, the response times are plotted as estimated by the model for the interactions of “condition” with “frequency” (left) and “TP” (right), with all other effects (of control and random variables) held at their mean.¹¹ Confidence bands refer to 1.96 standard errors of the effect of “frequency” or “TP”, respectively.

¹¹ We used the R packages *ggplot2* (Wickham 2016) and *schoenberg* (Simpson 2018) for these plots. By holding other effects at their mean, this visualization does not commit to any level for

Figure 4: Estimated response times by “condition” and “log frequency” (left)/“TP” (right).

The left panel shows that, expectedly, reduced forms produce slower responses across the board. The curves for the two conditions follow parallel developments up to a point, showing a facilitating effect of frequency. At higher frequencies, this trend continues with full forms while responses to reduced forms are somewhat delayed. A “chunking effect” of high frequency is thus observed for reduced forms only, although this effect is not very clear (with $p = 0.055$ for the smooth term, see Table 2). What is clear, however, is that the gap between the full and reduced variants widens at high frequencies.

The right panel shows again that recognition of reduced forms profits from higher TP, while full forms are not affected. The two linear curves overlap at high TPs, where the difference is almost levelled; it seems that reduction does not, or scarcely, delay recognition when the item is highly predictable from the immediate context.

the categorical control variables. Rather, the mean of the coefficients of the levels was used for each variable, as a way of factoring out the variable’s effect. This approach differs from available plotting functions in R packages such as *visreg* (Breheny and Burchett 2017) and *itsadug* (Van Rij et al. 2017), which set control variables to the reference level or the most frequent level. Setting a level affects the estimates and the size of the difference between the conditions, particularly where interactions pertain (such as *plosive_cluster* * *condition* in the present case). Therefore, the mean effects are the most neutral representation possible.

3.2.2 Effect of control variables

Since our original hypothesis concerned frequency and conditional probability, other properties of the input items are considered control variables. Yet, these control variables turn out to play an important role. Controlling for age reveals that older participants show somewhat slower responses, mainly on the full forms (upper left panel in Figure 5). It looks as though they did the task a little more cautiously rather than reacting quickly (they also responded slower to control items); note though, that the “older” end of the scale is represented only by few participants. Regarding gender, male participants show a greater delay with reduced forms (upper right panel). This effect is weak in the model ($t = 1.89$, $p = 0.06$), and while it has to be acknowledged, any far-reaching interpretation would seem out of place. In sum, there is a tendency for younger and male participants, respectively, to show a greater difference in their responses to full and reduced.

Figure 5: Response times to full and reduced items by control variables.

The other control factors that remain in the model (making a significant contribution to explaining the results) are those that measure properties of the verb preceding *to*, i.e., its duration and the phonological segments at its boundary. Responses to *to* are faster when the preceding verb has a longer duration, regardless of condition (lower left panel in Figure 5). A longer verb provides a longer span for its processing; this prompts a faster recognition of the upcoming item. Finally, when the final segment of the verb preceding *to* is an alveolar plosive (e.g., *need to*, *hate to*), the cluster ([dt] or [tt]) gets merged into a single segment – this case inhibits recognition of a reduced *to* but has no effect on full forms (lower right panel). The merger produces an effect of coalescence as it obscures a word boundary. This effect is discussed further in Section 4.3.

4 Discussion

There is a common (and common-sense) assumption that speakers aim for economy of effort and ease of articulation, leading to phonetic reduction, whereas listeners require explicitness and clarity, hence favoring full forms and clearly marked boundaries (cf. Lindblom 1990; Beckner et al. 2009: 16). This observation is confirmed by the finding that recognition of reduced forms of *to* generally takes longer than for full forms. However, reduction need not always cause problems for the listener. In the following, we summarize the results with regard to frequency information (surface frequency and transitional probability, Section 4.1) and the role of phonological context (4.2); Section 4.3 discusses the ramifications of our findings for the notions of speech perception and mental representation in cognitive linguistics.

4.1 Frequency information

The main question we asked was how frequency information interacts with reduction in speech perception – do listeners access and use this information in coping with reduced input?

For the surface frequency of the sequence (*V + to*), we hypothesized that high frequencies would help the recognition of reduced forms, since frequent sequences form an entrenched routine, and reduction is associated with frequency in speech production; we further expected to see a chunking effect delaying recognition of both full and reduced forms at very high frequencies (cf. Sosa and MacFarlane 2002; Kapatsinski and Radicke 2009).

The results we obtain for string frequency only partially match these expectations (see Figure 4, left panel). The linear effect of frequency on full forms (faster responses to items of higher frequency) suggests that there is a general facilitative role of string frequency in word recognition. This would imply that frequency simply serves as a base probability (cf. Hall et al. 2018: 2) – less frequent items are less expected, more frequent items are more expected – irrespective of the items' compositionality and their potential chunking. What frequency facilitates, then, is the processing of separate items in a sequence. Recognition of reduced forms, however, is aided by frequency only up to a point (around $\log \text{frequency} = 3$). If there is a frequency range in which listeners may expect reduction and therefore identify a reduced item more easily, this would be at these mid-high frequencies. Higher frequencies facilitate recognition only of the full forms but provide no further advantage with reduced forms. This suggests that if chunking interferes with accessing the element *to* (and offsets the facilitating frequency effect), it does so only when *to* is reduced. In other words, the activation of a holistic representation is based not only on frequency, but also on the phonetic form of the input.

For conditional probability, we expected that when *to* is highly probable (given the previous word), it will be recognized faster in general and perhaps especially aid recognition of reduced items. However, the results (Figure 4, right panel) show that recognition of full forms is insensitive to TP, suggesting that a higher probability has no general facilitating effect. Reduced forms, in contrast, show a strong linear effect, with higher TP continuously facilitating recognition. As transitional probability refers to predictability from context, it seems that listeners benefit from this information when it is most useful, in this case to recover a reduced form. At low TPs, *to* is least predictable from context, and it is here that reduction causes the greatest difficulty. At high TPs, *to* is more easily recognized in spite of reduction. If listeners draw on probability to derive a strong expectation of an upcoming *to*, they need less information from the input to identify it. On the other hand, when the full phonetic cue is available (i.e., full forms), it seems that listeners do not need to resort to the information from transitional probability as a further support in recognizing the item. One might suspect that the lack of an effect of TP with full forms is due to a floor effect on response times: when the signal is very clear (i.e., full forms), there is simply less room for listener expectations to improve recognition. However, responses to full forms are not consistently "fast", as can be seen from the strong effect of frequency on full forms (where response times to low frequency items are at ca. 600–750 ms, high frequency at 400–500 ms). This variance would, in principle, allow for an effect of TP as well, but this is not found.

It is noteworthy that we find strong effects of forward transitional probability (the probability of *to* given the preceding word), but not of backward transitional probability (the probability of *to* given the following word) or the surface frequency of the *to*-V_{inf} sequence. This may be due to the restricted set of fairly common verbs following *to* in our stimuli, which were designed to focus on the preceding rather than the following verb. In speech production, several corpus studies find measures involving the following word to have an impact, e.g., on duration and vowel reduction in function words such as *to* (Jurafsky et al. 2001; Bell et al. 2003). However, Bell et al. (2009) report this effect to hold only in mid- and low-frequency function words, whereas high-frequency items (such as *to*) are conditioned only by forward probability. They propose that high-frequency function words “often include alternate forms lacking onsets for the words with obstruent onsets (*the*, *that*, *to*). This makes them vulnerable to reduction when closely associated with previous words” (Bell et al. 2009: 108). Since our design included lenition of the /t/ onset of *to*, it appears that this association with previous words in production translates to higher expectations of reduction in perception. We cannot make a conclusive statement about any similar association with the following word.

4.2 Effect of phonological context

In addition to the effects of frequency and transitional probability, we observe a strong influence of phonological properties of the verb preceding *to*.

A longer duration of the verb accelerates recognition of a following *to* (see Figure 5, lower left panel). It seems that this is merely a side-effect, as longer words (e.g., *remember*) can be recognized before they are completed and thus allow listeners to more quickly predict and process the next item.

The inhibitory effect of reduction is much greater when /t/ merges with a preceding alveolar plosive (e.g., in *need to*, *pretend to*; Figure 5, lower right panel). Given that listeners are sensitive to fine-grained phonetic cues to identify word boundaries (cf. Fernandes et al. 2007), we can interpret this as an effect of on-line phonetic coalescence: when such a merged plosive is lenited, it marks a reduction not only at the onset of *to* but also at the offset of the preceding verb and thus blurs the word boundary. This creates a stronger connection (coalescence) between the verb and *to*, and may lead listeners to a non-compositional access path. The delay in recognition is then caused by the need to decompose the unit (V + *to*) in order to identify the element *to*. It would seem that this coalescence in phonetic realization has a similar effect to frequency-based chunking in fusing two elements together. Indeed, in natural speech production,

phonetic coalescence and frequency-based chunking are probably interconnected in the production-perception loop: a coalesced realization (e.g., “needa”) will reinforce a non-compositional mental representation, which will in turn increase the frequency of reduced realizations. Thus, a reduced merger of alveolar plosives seems favorable to the propagation of a single-unit representation of a high frequency sequence (*need to*).¹²

However, frequency-based chunking is seen as holistic storage in memory, so that the “chunk” can be accessed as an item, whereas on-line phonetic coalescence is a matter of articulation in speech production and needs to be decoded on the spot. The delayed responses to strongly coalesced realizations (reduction in the merged plosive, e.g., “needa”, “pretenda”) thus point to the gradient nature of chunking (cf. Ellis 2002). The fusion of two elements is represented in memory to varying degrees, and its activation depends on the form of the input. When the word onset is not clearly identifiable, listeners have greater difficulties in singling out the word *to* as separate from the previous word. This finding lines up with accounts that word-initial material has special importance in word recognition (Marslen-Wilson and Tyler 1980; McQueen et al. 2003; Astheimer and Sanders 2011; but see Connine et al. 1993).

4.3 Prediction, entrenchment, chunking

Our broad finding has been that reduction delays the recognition of an item, that frequency information can aid recognition, and that forward transitional probability is most clearly instrumental in recovering reduced forms. We also find a modest “chunking effect” of high frequency, albeit only on reduced forms. Thus, the activation of a chunk – which might interfere with listeners’ recognition of an individual element – depends not only on frequency but also on the phonetic form of the input (i.e., reduced). Moreover, reduction that obscures the word boundary (“plosive cluster” in our results) inhibits recognition, suggesting a connection between chunking and on-line phonetic coalescence. We now turn to a discussion of these results in light of relevant concepts of cognitive linguistics, in particular prediction, entrenchment and chunking.

¹² The reduction of /t/ in a plosive cluster comes naturally in connected speech and yet has the effect of blurring the word boundary, thus clearly marking coalescence. It might then be an important factor pushing for the conventionalization of coalesced variants (e.g., “needa”). In contrast, reduction in other phonological contexts is not as favorable to coalescence, since a reduced /t/ still marks the onset of *to* (*have to* [rə] or *trying to* [rə]; cf. Lorenz and Tizón-Couto 2017: 22). This might actually slow down the process of conventionalization of the contracted variant (“haveda” / “trynda”).

Frequencies of multi-word sequences are known to affect speech production (Arnon and Cohen Priva 2013; Gradoville 2017) – our results confirm that frequency also plays a role in speech perception (see also Van de Ven and Ernestus 2018), mainly in that low bigram frequency causes greater difficulty for word recognition. However, speech perception may also rely on predictability and probabilistic cues. Listeners may intuitively employ probabilistic measures related to the previous input in order to form expectations and anticipate the next word. Moreover, they draw on this information especially when the present input is difficult to recognize by itself (Ernestus et al. 2002; Mattys et al. 2012), so that prediction could be seen as a compensation strategy (cf. Pickering and Garrod 2007).

The present results suggest that frequency-based probabilities (derived from surface frequency and transitional probability) are available to listeners as an aid for anticipating upcoming words. This requires, minimally, a tacit knowledge of such probabilities.¹³ While surface frequency shows an effect with both full and reduced inputs, transitional probability is relied on more in the case of reduction, i.e., when phonetic cues are weaker. This is in line with other accounts of probabilistic prediction. Firstly, there is evidence in speech processing that conditional probabilities are less important than other, more direct cues (lexical information, phonological matching; Mattys et al. 2005; Franco and Destrebecqz 2012). Huettig and Mani (2016: 19) argue that probabilistic prediction “provides a ‘helping hand’ but is not necessary for language processing”. Secondly, the degree to which listeners reach for this “helping hand” is dependent on its utility given the task and goals at hand (Kuperberg and Jaeger 2016: 44–45). In a word monitoring task, this utility is presumably very high at any point. The results of the present experiment suggest that the utility of (probabilistic) frequency information is especially relevant when the acoustic information is less complete (as in reduction).

Another current issue in cognitive linguistics concerns the mental representation of multi-word sequences, namely the tension between holistic storage (chunking) and procedure strengthening as two different possible interpretations of the notion of entrenchment: are highly frequent phrases stored as single units (Bybee 2006; Diessel 2007; Ellis et al. 2009), or are their pieces more rapidly

13 The role of prediction in a word-monitoring task may, of course, be different from natural speech comprehension. Firstly, the experimental task demands specific attention to form rather than meaning; consequently, the predictive measures we consider are purely based on surface forms, not on semantic or situational information. Secondly, the set-up is “prediction-encouraging” (cf. Huettig and Mani 2016: 26) – in monitoring for a given word, participants may pick up on any cue that helps them assess the likelihood of the target word coming up.

assembled due to continual usage (Divjak and Caldwell-Harris 2015: 66–67)? Speech-monitoring studies have provided partial evidence for faster retrieval of elements from more frequent bigrams, which is offset by a chunking effect at (very) high bigram frequency (Kapatsinski and Radicke 2009). The present study contributes to this research by putting reduction into the picture, which differentiates these effects. On the one hand, procedure strengthening is seen in the continuous facilitating effect of frequency on full forms and the capacity of transitional probability to balance out the adverse effect of reduction. On the other hand, a modest chunking effect of frequency is found for reduced items, in the delayed response times at high frequencies.

Procedure strengthening and chunking may be conceived of as stages of a continuous process (cf. Langacker 1987: 59–60; Blumenthal-Dramé 2012: 68–69, 104). With increasing frequency, a sequence is increasingly entrenched as a procedure, and with increasing entrenchment, boundaries between the individual components become less important. The end result is a chunk, a mental representation as a single unit. However, the present results suggest that even with very high frequency, chunking is not inevitable. Listeners may perceive a bigram as either a chunked item or a compositional sequence, i.e., they have two access paths available, that of a memorized chunk and that of a compositional sequence. The activation of one or the other access path is affected not just by frequency but by the properties of the input signal. When reduced forms occur in a high-frequency bigram (e.g., “*needa*” *need to*, “*havda*” *have to*), they lead the listener into perceiving a chunked item. Full forms on the other hand encourage a compositional access. In short, for high-frequency bigrams, the presence or absence of reduction determines whether listeners perceive a chunked item or a compositional sequence.

This conclusion is not compatible with a one-dimensional concept of entrenchment, where at some point of frequency a sequence becomes reanalyzed as a chunk. Rather, we need a “continuous” approach in which “the difference between more and less frequent is [...] one of degree, rather than specifying whether the sequence is stored vs. computed” (Caldwell-Harris et al. 2012: 3–4), and, crucially, a concept of chunking as “global precedence”: “a configuration of elements qualifies as a cognitive chunk, if the configuration as a whole is cognitively more prominent than its individual component parts” (Blumenthal-Dramé 2018: 138). In this view, there is a mental representation of the “whole”, i.e., a stored chunk, but what is activated in perception – whole or parts – is a matter of prominence. “Cognitive prominence” may be seen as a kind of base activation of the holistic representation; but to explain the discrepancy in the frequency effects on full and reduced forms, we must add a perceptual prominence. With reduction, the individual parts are made less

prominent, so a perception of the “whole” – if accessible as a mental representation – takes over. With full forms, the individual items are clearly distinguishable and can therefore be accessed without interference from the holistic representation.

The finding that the chunking effect is limited to reduced forms, may also be interpreted in terms of stored pronunciation variants. For single words, listeners are responsive to pronunciation variants, and even to the correlation between word frequency and phonetic reduction (Ranbom and Connine 2007; Bürki et al. 2011; Mitterer and Russell 2013; Brand and Ernestus 2018). If high-frequency sequences (e.g., *need to*) are liable to phonetic reduction (e.g., “needa”), these reduced forms may also be stored as variants. In the experiment, it is possible that reduction of *to* in very high frequency “V-*to*” sequences activates these variant representations. These would be variants of a single unit in which the original elements are less prominent. The mental representation of a chunked item will then delay the recognition of a reduced element (in this case *to*) because it also includes reduced variant forms (e.g., *needa*).

Finally, there is the special case where reduction occurs at a merged word boundary (e.g., “needa” *need to*, but also “hayda” *hate to*). The observed recognition delay suggests that listeners have to deal with an additional limitation that is somehow related to chunking, but probably not equivalent: in frequency-based chunking the processing difficulty is caused by the interference of an item stored in memory; however, a reduced word boundary can be a matter of on-line phonetic coalescence in speech production and simply needs to be decoded by the listener on the spot. Further research could elucidate the exact relation between on-line phonetic coalescence and memorized chunks.

5 Conclusion

The study addresses the question of how frequency information affects the recovery of reduced forms in speech perception. We have reported a word-monitoring experiment with full and reduced variants of the target item *to* in V-*to*-V_{inf} sequences in English. The results shed light on how listeners draw on probabilistic cues and access mental representations of linguistic structures.

The results confirm that language users have rich stochastic information available, and they relate it to the input they receive in speech perception.

However, this information is not applied in a broad across-the-board fashion. Rather, its effect depends in part on the strength of the other cues provided by the signal. The findings concur with a notion of word recognition as a process where contextual information is permanently available to listeners, but primacy is given to information in the signal at the target itself. Thus, what listeners hear (information from the signal) can easily override what they expect (contextual information) but not as easily vice versa: uncertainty in expectation (low predictability) should be less problematic to recognition than uncertainty in the signal; however, when listeners cannot be sure of what they are hearing (e.g., due to reduction), then the effect of contextual information remains high during recognition.

In terms of mental representations, the results support the notion that frequent collocations are stored in the mind as a single chunk, but this single-unit representation does not override access to the component parts of the sequence. We have argued that a holistic and a compositional representation are both available. Activation of the chunk rather than the compositional sequence is determined not only by entrenchment through frequency, but by articulatory cues from the input (reduction, and perhaps blurred word boundaries). The concept of a “chunk” that fits best with this finding is a perceptual one, where a holistic perception of a structure takes precedence over a compositional one. In this view, chunking is a matter of degree, but does imply the presence and entrenchment of a chunk as a single unit. Possibly, these chunks include a representation of reduced pronunciation variants that would in turn be activated in recognition.

Acknowledgements: We would like to thank the (Chief and Associate) editors of *Cognitive Linguistics* – John Newman, Dagmar Divjak, Laura Janda and Benedikt Szmrecsanyi – for their scrutiny, advice and valuable suggestions; and two anonymous reviewers for their constructive criticism and insightful comments. We would also like to acknowledge the useful feedback we received from colleagues at our institutes in Freiburg, Vigo and Rostock, and at several conferences. All of these have immensely improved this paper; any remaining obscurities are entirely our own responsibility.

We are grateful to the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (grant no. FFI2016-77018-P and grant no. IJCI-2015-25843) and Xunta de Galicia (grant no. ED431C 2017/50) for generous financial support; and to Wissenschaftliche Gesellschaft Freiburg for a grant for participant compensations.

Appendix A. List of experimental items

List of experimental items by surface frequency and transitional probability in the spoken section of COCA (Davies 2008):

Verb	Frequency (per 1mil)	Log Frequency	Verb	Transitional Probability
craved	0.03	−3.60	remember	0.005
deign	0.15	−1.92	mean	0.009
yearning	0.21	−1.56	fear	0.011
fear	0.29	−1.23	regret	0.036
regret	0.42	−0.87	forget	0.048
itching	0.42	−0.87	craved	0.059
desire	0.46	−0.78	happen	0.061
strive	0.56	−0.58	love	0.099
aspire	0.56	−0.58	dare	0.115
cease	0.66	−0.42	started	0.117
dare	1.18	0.16	offered	0.130
remember	1.50	0.40	prepared	0.135
volunteered	2.19	0.79	have	0.136
pretend	2.60	0.95	cease	0.169
vowed	3.29	1.19	yearning	0.215
forget	3.70	1.31	deserve	0.221
prefer	5.18	1.65	pretend	0.234
deserve	5.62	1.73	hate	0.237
offered	5.65	1.73	promised	0.254
promised	8.15	2.10	began	0.254
prepared	8.23	2.11	desire	0.287
mean	11.09	2.41	strive	0.300
intend	11.85	2.47	prefer	0.304
hate	11.90	2.48	used	0.356
plan	12.70	2.54	failed	0.365
chose	12.78	2.55	aspire	0.377
failed	17.15	2.84	agreed	0.392
happen	18.33	2.91	plan	0.400
refused	23.79	3.17	chose	0.417
agreed	24.60	3.20	volunteered	0.435
love	30.33	3.41	itching	0.469
started	33.73	3.52	continue	0.472
began	33.97	3.53	need	0.499
tend	35.94	3.58	seem	0.562

(continued)

(continued)

Verb	Frequency (per 1mil)	Log Frequency	Verb	Transitional Probability
seem	83.05	4.42	vowed	0.578
continue	91.57	4.52	want	0.648
used	114.41	4.74	going	0.732
need	292.39	5.68	refused	0.746
trying	464.55	6.14	intend	0.826
have	951.38	6.86	tend	0.889
want	1053.52	6.96	trying	0.901
going	2378.61	7.77	deign	0.941

References

Altmann, Gerry T. M. & Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73. 247–264.

Arnon, Inbal & Uriel Cohen Priva. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech* 56(3). 349–371.

Arnon, Inbal & Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62. 67–82.

Asstheimer, Lori B. & Lisa D. Sanders. 2011. Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia* 49(12). 3512–3516.

Baayen, R. Harald & Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2). 12–28.

Baayen, R. Harald, Shravan Vasishth, Reinhold Kliegl & Douglas Bates. 2017. The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language* 94. 206–234.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.

Barth, Danielle & Vsevolod Kapatsinski. 2017. A multimodel inference approach to categorical variant choice: Construction, priming and frequency effects on the choice between full and contracted forms of *am*, *are* and *is*. *Corpus Linguistics and Linguistic Theory* 13(2). 203–260.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann (The ‘Five Graces Group’). 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59(1). 1–26.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60. 92–111.

- Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113(2). 1001–1024.
- Blumenthal-Dramé, Alice. 2012. *Entrenchment in usage-based theories: What corpus data do and do not reveal about the mind*. Berlin: Mouton de Gruyter.
- Blumenthal-Dramé, Alice. 2018. Entrenchment from a psycholinguistic and neurolinguistic perspective. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning*, 129–152. Berlin: Mouton de Gruyter.
- Boersma, Paul & David Weenink. 2016. *Praat: Doing phonetics by computer* [computer program]. Version 6.0.14. <http://www.praat.org/> (accessed 2 February 2016).
- Brand, Sophie & Mirjam Ernestus. 2018. Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: Evidence from native listeners and learners of French. *The Quarterly Journal of Experimental Psychology* 71(5). 1240–1259.
- Breheny, Patrick & Woodrow Burchett. 2017. Visualization of regression models using visreg. *The R Journal* 9. 56–71.
- Brown, Meredith, Laura C. Dilley & Michael K. Tanenhaus. 2012. Real-time expectations based on context speech rate can cause words to appear or disappear. In Naomi Miyake, David Peebles & Richard P. Cooper (eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 1374–1379. Austin: Cognitive Science Society.
- Bürki, Audrey & Ulrich H. Frauenfelder. 2012. Producing and recognizing words with two pronunciation variants: Evidence from novel schwa words. *The Quarterly Journal of Experimental Psychology* 65(4). 796–824.
- Bürki, Audrey, Malte C. Viebahn, Isabelle Racine, Cassandre Mabut & Elsa Spinelli. 2018. Intrinsic advantage for canonical forms in spoken word recognition: Myth or reality? *Language, Cognition and Neuroscience* 33(4). 494–511.
- Bürki, Audrey, F., Xavier Alario & Ulrich H. Frauenfelder. 2011. Lexical representation of phonological variants: Evidence from pseudohomophone effects in different regiolects. *Journal of Memory and Language* 64. 424–442.
- Bushong, Wednesday & T. Florian Jaeger. 2017. Maintenance of perceptual information in speech perception. In Glenn Gunzelmann, Andrew Howes, Thora Tenbrink & Eddy J. Davelaar (eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 186–191. Austin: Cognitive Science Society.
- Bybee, Joan. 2002. Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition* 24(2). 215–221.
- Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.
- Caldwell-Harris, Catherine L., Jonathan Berant & Shimon Edelman. 2012. Entrenchment of phrases with perceptual identification, familiarity ratings, and corpus frequency statistics. In Dagmar Divjak & Stefan T. Gries (eds.), *Frequency effects in language representation*, 165–194. Berlin: Mouton de Gruyter.
- Connine, Cynthia M. 2004. It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review* 11(6). 1084–1089.

- Connine, Cynthia M., Dawn G. Blasko & Debra Titone. 1993. Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language* 32. 193–210.
- Connine, Cynthia M. & Eleni Pinnow. 2006. Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review* 23(3). 235–245.
- Connine, Cynthia M., Larissa J. Ranbom & David J. Patterson. 2008. Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics* 70(3). 403–411.
- Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA)*. 450 million words, 1990–present. <http://corpus.byu.edu/coca/> (accessed 1 April 2016).
- Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25(2). 108–127.
- Divjak, Dagmar & Catherine L. Caldwell-Harris. 2015. Frequency and entrenchment. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*, 53–75. Berlin: Mouton de Gruyter.
- Ellis, Nick. 2002. Frequency effects in language processing. *Studies in Second Language Acquisition* 24(2). 143–188.
- Ellis, Nick C., Eric Frey & Isaac Jalkanen. 2009. The psycholinguistic reality of collocation and semantic prosody (1): Lexical access. In Ute Römer & Rainer Schulze (eds.), *Exploring the lexis–Grammar interface*, 89–114. Amsterdam: John Benjamins.
- Ernestus, Mirjam & R. Harald Baayen. 2007. The comprehension of acoustically reduced morphologically complex words: The roles of deletion, duration and frequency of occurrence. In Jürgen Trouvain & William J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 773–776.
- Ernestus, Mirjam, R., Harald Baayen & Rob Schreuder. 2002. The recognition of reduced word forms. *Brain and Language* 81. 162–173.
- Fernandes, Tânia, Paulo Ventura & Régine Kolinsky. 2007. Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception & Psychophysics* 69(6). 856–864.
- Franco, Ana & Arnaud Destrebecqz. 2012. Chunking or not chunking? How do we find words in artificial language learning? *Advances in Cognitive Psychology* 8(2). 144–154.
- Frank, Stefan & Roel Willems. 2017. Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Language, Cognition and Neuroscience* 32(9). 1–12.
- Gahl, Susanne & Susan M. Garnsey. 2004. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language* 80(4). 748–775.
- Gradoville, Michael. 2017. The cognitive representation of multi-word sequences: A usage-based approach to the reduction of Fortalezense Portuguese *para*. *Lingua* 199. 94–116.
- Gregory, Michelle L., William D. Raymond, Alan Bell, Eric Fosler-Lussier & Daniel Jurafsky. 1999. The effects of collocational strength and contextual predictability in lexical production. *Communication and Linguistic Studies* 35. 151–166.
- Hall, Kathleen Currie, Elizabeth Hume, T. Florian Jaeger & Andrew Wedel. 2018. The role of predictability in shaping phonological patterns. *Linguistics Vanguard* 4(s2).
- Hartsuiker, Robert J. & Agnes Moors. 2018. On the automaticity of language processing. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning*, 201–226. Berlin: Mouton de Gruyter.

- Hope, Ryan M. 2013. *Rmisc: Ryan miscellaneous*. R package version 1.5. <https://CRAN.R-project.org/package=Rmisc>.
- Huetting, Falk & Nivedita Mani. 2016. Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience* 31(1). 19–31.
- Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 229–254. Amsterdam: John Benjamins.
- Kampstra, Peter. 2008. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software* 28(Code Snippet 1). 1–9. <http://www.jstatsoft.org/v28/c01/>.
- Kapatsinski, Vsevolod & Joshua Radicke. 2009. Frequency and the emergence of prefabs: Evidence from monitoring. In Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali & Kathleen Wheatley (eds.), *Formulaic language. Vol. II: Acquisition, loss, psychological reality, functional explanations* (Typological Studies in Language 83), 499–520. Amsterdam: John Benjamins.
- Krause, Florian & Oliver Lindemann. 2014. Expyriment: A Python library for cognitive and neuroscientific experiments. *Behavior Research Methods* 46(2). 416–428.
- Kuperberg, Gina R. & T. Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31(1). 32–59.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar. vol. I: Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. 2000. A dynamic usage-based model. In Michael Barlows & Suzanne Kemmer (eds.), *Usage based models of language*, 1–63. Stanford: CSLI Publications.
- Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H and H theory. In William J. Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer Academic Publishers.
- Lorenz, David. 2013. *Contractions of English semi-modals: The emancipating effect of frequency*. NIHIN Studies. Freiburg: Universitätsbibliothek Freiburg.
- Lorenz, David & David Tizón-Couto. 2017. Coalescence and contraction of V-to-V_{inf} sequences in American English – Evidence from spoken language. *Corpus Linguistics and Linguistic Theory*. Advance online publication. <https://doi.org/10.1515/cllt-2015-0067>.
- Marslen-Wilson, William & Lorraine K. Tyler. 1980. The temporal structure of spoken language understanding. *Cognition* 8(1). 1–71.
- Mathôt, Sebastiaan, Daniel Schreij & Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44(2). 314–324.
- Mattys, Sven L., Matthew H. Davis, Ann R. Bradlow & Sophie K. Scott. 2012. Speech recognition in adverse conditions: A review. *Language and Cognitive Processes* 27(7/8). 953–978.
- Mattys, Sven L., Laurence White & James F. Melhorn. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General* 134(4). 477–500.
- McDonald, Scott A. & Richard C. Shillcock. 2003. Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science* 14(6). 648–652.
- McQueen, James M., Delphine Dahan & Anne Cutler. 2003. Continuity and gradedness in speech processing. In Niels O. Schiller & Antje S. Meyer (eds.), *Phonetics and phonology in language comprehension and production*, 39–78. Berlin: Mouton de Gruyter.

- Mitterer, Holger & Kevin Russell. 2013. How phonological reductions sometimes help the listener. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(3). 977–984.
- Patterson, David, Paul C. LoCasto & Cynthia M. Connine. 2003. Corpora analyses of schwa deletion in conversational American English. *Phonetica* 60(1). 45–69.
- Pickering, Martin J. & Simon Garrod. 2007. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences* 11(3). 105–110.
- Pitt, Mark A. 2009. The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology: Human Perception and Performance* 35(3). 896–910.
- Pitt, Mark A., Laura Dilley & Michael Tat. 2011. Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics* 39(3). 304–311.
- R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Racine, Isabelle, Audrey Bürki & Elsa Spinelli. 2014. The implication of spelling and frequency in the recognition of phonological variants: Evidence from pre-readers and readers. *Language, Cognition and Neuroscience* 29(7). 893–898.
- Ranbom, Larissa. J. & Cynthia M. Connine. 2007. Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language* 57(2). 273–298.
- Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133(1). 140–155.
- Simpson, Gavin L. 2018. *schoenberg: ggplot-based graphics and other useful functions for GAMs fitted using mgcv*. R package version 0.0-6. <https://github.com/gavinsimpson/schoenberg>
- Simpson, Greg B., Robert R. Peterson, Mark A. Casteel & Curt Burgess. 1989. Lexical and sentence context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(1). 88–97.
- Sosa, Anna Vogel & James MacFarlane. 2002. Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word *of*. *Brain and Language* 83(2). 227–236.
- Tizón-Couto, David & David Lorenz. 2018. Realizations and variants of *have to*: What corpora can tell us about usage-based experience. *Corpora* 13(3). 371–392.
- Tremblay, Antoine & Benjamin V. Tucker. 2011. The effects of n-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon* 6(2). 302–324.
- Tucker, Benjamin V. 2011. The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics* 39(3). 312–318.
- Van Berkum, Jos J. A., Colin M. Brown, Pienie Zwitserlood, Valesca Kooijman & Peter Hagoort. 2005. Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(3). 443–467.
- Van de Ven, Marco & Mirjam Ernestus. 2018. The role of segmental and durational cues in the processing of reduced words. *Language and Speech* 61(3). 358–383.
- Van Rij, Jacolien, Martijn Wieling, R. Harald Baayen & Hedderik Van Rijn. 2017. *itsadug: Interpreting time series and autocorrelated data using GAMMs*. R Package Version 2.3.
- Warner, Natasha & Benjamin V. Tucker. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America* 130(3). 1606–1617.
- Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer.
- Wood, Simon N. 2006. *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman and Hall/CRC Press.

- Wood, Simon N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73(1). 3–36.
- Wright, Daniel B. & Kamala London. 2009. *Modern regression techniques using R: A practical guide*. London: SAGE.
- Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cog-2017-0138>; the data are available at <https://doi.org/10.18710/7TSABU>).