



HHS Public Access

Author manuscript

Lang Cogn Neurosci. Author manuscript; available in PMC 2016 January 01.

Published in final edited form as:

Lang Cogn Neurosci. 2015 ; 30(1-2): 88–102. doi:10.1080/01690965.2013.840733.

Synthesizing meaning and processing approaches to prosody: performance matters

Jennifer E. Arnold and

University of North Carolina, Chapel Hill, NC

Duane G. Watson

University of Illinois at Urbana-Champaign

Abstract

Words vary in acoustic prominence; for example repeated words tend to be reduced, while focused elements tend to be acoustically prominent. We discuss two approaches to this phenomenon. On the message-based view, acoustic choices signal the speaker's meaning or pragmatics, or are guided by syntactic structure. On the facilitation-based view, reduced forms reflect facilitation of production processing mechanisms. We argue that message-based constraints correlate systematically with production facilitation. Moreover, we argue that discourse effects on acoustic reduction may be at least partially mediated by processing facilitation. Thus, research needs to simultaneously consider both competence (message) and performance (processing) constraints on prosody, specifically in terms of the psychological mechanisms underlying acoustic reduction. To facilitate this goal, we present preliminary processing models of message-based and facilitation-based approaches, and outline directions for future research.

Keywords

prosody; acoustic prominence; reference; language production; audience design

INTRODUCTION

A common approach to understanding prosody assumes that speakers choose prosodic forms to reflect some aspect of their meaning; that is, on the basis of their knowledge of the grammatical and pragmatic rules of their language. This view further assumes that processing mechanisms, otherwise known as performance constraints, are irrelevant to accounts of the underlying prosodic representations. In this paper, we argue that this distinction limits the field's ability to progress. Our argument is rooted in our goal of understanding the psychological mechanisms that drive acoustic variation in language.

We take one prosodic phenomenon, acoustic prominence, and compare two types of approaches that have been taken towards understanding the distribution of prominent words: one based on linguistic competence (the “message-based approach”) and the other based on

linguistic performance (the “facilitation-based approach”). We argue that synthesizing the work from these two approaches will result in greater progress in understanding the mechanisms that underlie prosodic variation. As a first step towards this goal, we propose two general models for acoustic prominence, and outline directions for future research.

Explanations of Acoustic Prominence

Words vary in acoustic prominence, ranging from highly prominent forms (*you ate WHAT?*) to reduced ones (*what I ate was a BAGEL*). In English, this contrast between reduced and prominent forms is achieved by variation in duration, pitch, pitch movement, and amplitude (Ladd, 2008). In many cases acoustic prominence is linked to the use of an accented form, as opposed to an unaccented one. However, even within accent categories there is variation in acoustic prominence (Breen, Fedorenko, Wagner, & Gibson 2010; Watson & Arnold, 2005).

Why does speech vary in this way? We consider two classes of explanation in this paper: a message-based approach and a facilitation-based approach.

The message-based approach refers to explanations of variation in linguistic form in terms of the speaker’s meaning or the function of the utterance. Put another way, people say things a particular way because the grammar selects that form for their intended message, whether at the syntactic, semantic, or pragmatic level. In this paper we focus on how acoustic prominence varies as a function of pragmatic appropriateness. For example, words are typically reduced after they have been mentioned, e.g. *I like BAGELS. You like bagels too*, or when they are predictable in context (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Brown, 1983; Fowler & Housum, 1987; Jurafsky, et al., 2001). An assumption of the message-based approach is that speakers select a level of prominence in order to appropriately mark its information status. As a generalization, given information is marked with a reduced form, while new or contrastive information is acoustically prominent (Halliday, 1967).

A second class of explanation focuses on the fact that variation in acoustic prominence reflects the relative difficulty of production. For example, longer word duration correlates with disfluency (Bell et al. 2003; Clark and Fox Tree, 2002) and difficult production circumstances (Ferreira & Swets, 2002). Conceptual difficulty also can induce higher pitch (Christodoulou, 2009). Conversely, conceptual facilitation results in shorter durations (Balota, Boland, & Shields, 1989). These findings suggest that acoustic reduction can also result from facilitation within the production processing system.

The distinction between the message and processing accounts of prosody partly reflects a sharp distinction drawn by linguistic theories between linguistic competence and linguistic performance. ‘Competence’ refers to the knowledge we have about linguistic elements and the algorithms for combining them grammatically into new words and sentence structures. In contrast, ‘performance’ refers to the real-time use of this knowledge, and the cognitive and physical systems that must be engaged to do so. This distinction is usually drawn for the purpose of highlighting the importance of understanding linguistic competence as a window onto the cognitive architecture underlying human language abilities (e.g., Chomsky, 1965). Performance factors are often considered, at best, irrelevant to the question of understanding

how language is represented in the human mind or, at worst, noise that interferes with the analysis of a native speaker's linguistic knowledge.

The core problem with this distinction in prosody is that competence and performance constraints are highly correlated. This means that acoustic patterns cannot be assumed to reflect purely one or the other without careful consideration of both. We illustrate this argument by reviewing research on the relation between information status and variation between acoustically prominent and reduced pronunciations. While prosody also reflects other constraints, like syntactic structure or lexical stress, these are outside the scope of this paper. We demonstrate that some findings can be explained by both the message-based and production facilitation accounts, and moreover that these accounts are at least partially confounded. We propose that research on production mechanisms needs to consider both message-based and facilitation mechanisms, particularly because message-based effects may be mediated by a facilitation mechanism. We present some preliminary ideas of what these mechanisms may look like, and suggest directions for future research.

INFORMATION STATUS AND ACOUSTIC PROMINENCE

It is well established that words vary in their information status, which correlates with acoustic prominence or reduction. We discuss evidence for this correlation in terms of two characterizations of information structure: discourse status/focus marking, and predictability. We also consider the role of audience design. We then consider how each of these is accounted for by message-based and facilitation-based theories.

Discourse Status effects

Most scholars recognize a contrast between given and new information. In *I like BAGELS. You like bagels too*, the first mention of bagels is considered new, whereas the second is given. Given status is most frequently defined in terms of linguistic mention, but theories of information status acknowledge that information can be treated as given if it is evoked in other ways, such as being visually available (Chafe, 1987, 1994; Clark & Marshall, 1981; Prince, 1981), or inferrable from other given information (Prince, 1992; Schwartz child, 1999). Intuitively, it is more felicitous to produce a word prominently on the first mention, whereas subsequent mentions seem more natural with a more reduced pronunciation.

This intuition is borne out in the empirical literature. Fowler and Housum (1987; see also Fowler, 1988) examined the duration of first and second mentions in speech, and found that second mentions were shorter and less intelligible. Similarly Bard and colleagues (Bard et al., 2000; Bard & Aylett, 2004) examined speech in the Map task corpus, and found that first mentions were both longer and more intelligible than subsequent mentions. Intelligibility indexes a listener's ability to understand a word when it has been removed from its context and presented in isolation; those that are phonetically reduced carry less information, and are thus less intelligible.

The simple contrast between given and new provides a rough but robust categorization that maps onto pronunciation variation. However, a full characterization of information status requires finer-grained distinctions. Repeated mention of an object is most likely to be

reduced when the second mention occurs in the same syntactic position as the first one, e.g. *The ball touches the cone*; *The ball touches the star*, as opposed to *The cone touches the ball*; *The ball touches the star* (Terken & Hirschberg, 1994; see also Watson & Arnold, 2005; Watson, 2010). Parallelism effects may be related to other contrastive effects on prosody, e.g. the tendency to accent contrastive information (*Not the RED ball, the BLUE ball*; see Wagner & Klassen, under review).

There is also a great deal of work investigating how prominence interacts with other levels of linguistic representation such as semantics and syntax. For example, some researchers have worked on formalizing the relationship between the distribution of accents in a sentence and information structure, as well as understanding the syntactic and semantic constraints that influence prominence placement (e.g. Schwarzschild; 1999; Gussenhoven, 1983; Selkirk, 1996). More recently, Breen, et al. (2010) tested the extent to which speakers prosodically marked different types of focus and contrastiveness in sentences like *Damon fried an omelet*, under different information status conditions. In three experiments, they found that speakers marked focused information with longer durations, larger f0 excursions, longer following pauses, and greater intensity, compared to given information. Furthermore, speakers reliably marked focus location (Damon vs. fried vs. omelet) and focus breadth, and in Experiments 2 and 3 (but not Exp. 1) they also marked focus type (contrastive vs. noncontrastive). In keeping with the message-based account, listeners (in a perception experiment) were able to distinguish focus location extremely well, and were moderately able to distinguish focus type. However, listeners were not able to distinguish wide (*What happened this morning?*) and narrow focus (*What did Damon fry?*).

The correlation between accenting and discourse status also guides comprehension. Listeners prefer to interpret reduced forms as referring to given information, and acoustically prominent forms as referring to new information (Arnold, 2008b; Bock & Mazella, 1983; Dahan, Tanenhaus, & Chambers, 2002). For example, participants in Arnold's (2008b) experiment were eye-tracked while they followed simple instructions like *Put the bacon on the circle. Now put the bagel/BAGEL on the square*. When the target word bagel was unaccented, listeners initially looked at the given object (the bacon), but not when it was accented. This finding echoed the results of Dahan et al., (2002), who also found that accented tokens elicited a bias toward information that was given but not highly focused.

Predictability

Some theories about information status suggest that givenness is at least partly defined in terms of predictability (e.g., Prince, 1981). This definition predicts that information that is retrievable based on the linguistic or discourse context can be produced with greater acoustic reduction.

Indeed, corpus analyses have found that when a word is predictable in context, it tends to be shorter, and undergo segment deletion. Word predictability can be conditioned on the frequency of co-occurrence with other words (Bell et al., 2009; Jurafsky, Bell, Gregory, & Raymond 2001) or semantic content (Gregory et al., 2000). Words that are frequent or probable in context tend to be reduced (Bell et al., 2009; Jurafsky et al., 2001). Word probabilities have been shown to affect reduction at acoustic, lexical, phrasal, and syntactic

levels, which has led to theories that acoustic reduction is constrained by pressures to produce information at a rate that results in uniform information density (Aylett & Turk, 2004; Jaeger, 2006, 2010; Levy & Jaeger, 2008).

Aside from word predictability, predictability at other levels of representation can influence acoustic variation. The likelihood of syntactic structures themselves can modulate fluency and word pronunciation, with shorter pronunciations in more frequent structures than less frequent ones (Gahl & Garnsey, 2004; Tily et al. 2009),

In addition, the predictability of references themselves can affect acoustic reduction.. This referential predictability refers to the likelihood that a particular referent will be mentioned, and not the probability of the word itself, although it is likely to occur in similar contexts. For example, when referents are probable in context, they are likely to have greater activation, thus facilitating message planning and formulation. Such referential predictability effects were observed in a verbal game of Tic Tac Toe (Watson, Arnold, & Tanenhaus, 2008). When the game constraints made a move predictable, the utterance was shorter in duration than the unpredictable moves (see also Kahn & Arnold, 2012; Lam & Watson, 2010). Referential predictability is also associated with repeated mention, in that speakers tend to re-mention things that have occurred recently, especially in syntactically or semantically prominent positions (Arnold, 1998, 2010; Grosz, Weinstein, & Joshi, 1995; Givon, 1983).

Effects of common ground and audience design

An additional question is whether acoustic prominence is sensitive to the speaker's estimation of the addressee's knowledge or attention. This issue is orthogonal to questions about whether information status affects acoustic variation, in that it is conceivable that discourse status, focus, and predictability effects are all calculated on the basis of the speaker's knowledge of the current discourse context, or based on assumptions about the listener's knowledge due to linguistic co-presence (Clark & Marshall, 1981).

Nevertheless, proposals of information status frequently specify that acoustic prominence variation is only sensitive to information distinctions that are assumed to be in common ground. That is, for information to be given, it must be assumed to be known to all discourse participants (Gundel, Hedberg, & Zacharaski, 1993; Chafe, 1994). For example, Baumann and Grice (2006) suggest that a pitch accent is used to mark either a) the degree of activation of the referent in "the assumed (immediate) consciousness of the listener," (p. 6), or b) the speaker's wish to highlight information as noteworthy (see also Baumann & Hadelich, 2003). Another idea is that speakers may use prominent forms when the addressee's attention is not already on the referent, because more explicit bottom-up input is needed to facilitate processing (Rosa, Finch, Bergeson, & Arnold, 2013, this volume).

Questions about whether acoustic prominence is sensitive to audience design have been extensively discussed in the literature (Arnold, 2008a; Arnold, Kahn, & Pancani, 2012; Bard et al., 2000; Bard & Aylett, 2004; Brennan & Hanna, 2009; Gahl, Yao, & Johnson, 2012; Galati & Brennan, 2010; Kahn & Arnold, under review, this volume; Rosa et al., 2013, this volume). Although audience design questions are orthogonal to information status effects,

later we will consider how addressee-oriented effects may also reflect speaker-internal processing constraints.

In sum, there is a well-known correlation between acoustic prominence and information status, which can be characterized in terms of discourse status/focus structure, predictability, and common ground. It is likely that these effects are not independent of each other, in that given/salient discourse status correlates with predictability, and new discourse status often co-occurs with utterance focus. The key point here is that the contrast between acoustically reduced and acoustically prominent tokens varies as a function of information status. The question is, why does this relationship exist? We consider two broad theoretical approaches to this problem.

THEORETICAL APPROACHES: MESSAGE VS. FACILITATION APPROACHES

The message-based approach to explaining acoustic prominence

Many linguistic theories explain acoustic prominence in terms of contextually-based form selection: speakers select appropriate acoustic forms on the basis of semantic or pragmatic rules in their language, which call on information structure representations that define the conditions for each rule. A simplistic rule would be for words with “new” information status to be more prominent than words with a “given” status. This view underlies claims that, for example “deaccenting [is a] marker for given information” (Baumann & Hadelich, 2006). On this view, the context selects for (or licenses) a particular form, and speakers’ ability to use this form stems from their knowledge of the pragmatic norms of English. The core idea to this approach is that variation in linguistic form – e.g., a particular accent, such as H*, or perhaps acoustic prominence in and of itself -- reflects the communicative function of language. Put another way, this view suggests that linguistic forms reflect the speaker’s meaning – or, by extension, they reflect the grammatical and pragmatic rules about form that allow listeners to derive the speaker’s message from the linguistic input.

This “message-based” view encompasses numerous theoretical approaches to acoustic prominence, including both formal and functional theories. The purpose of this section is not to evaluate any particular proposal, nor to provide a comprehensive review. Instead, we mention a few as examples of this approach.

A number of authors have proposed explanations of accenting in terms of formal grammatical rules relating meaning to intonational form. One such view comes from autosegmental theories of accent placement in languages like English. For example, Pierrehumbert and Hirschberg (1990) propose a grammar of intonation for English, in which pitch accents, phrase accents, and boundary tones are used combinatorily to encode aspects of the utterance meaning. Part of this system includes marking information status, for example the H* accent is proposed to indicate that information is discourse-new (p. 289).

Similarly, some theories assume that an abstract focus structure serves as an intermediary between prosodic structure and information structure, where focus structure is part of a particular language’s grammar (e.g. Schwarzschild; 1999; Gussenhoven, 1983; Selkirk, 1996,

Wagner & Klassen; under review). Steedman (2000) has argued that pitch accents, along with other aspects of prosody like intonational boundaries, are part of the semantic and syntactic representations of a sentence.

Importantly, theories about prosodic structure are not intended as psychological processing mechanisms. Moreover, some theories specify the conditions under which particular forms are licensed, rather than the preferred selection (e.g., Grosz et al. 1995; Gundel et al., 1993)¹. However, we contend that the next major step in the field is to integrate what is known about prosodic form with a psychologically plausible mechanism that explains how speakers choose forms on any given occasion. In doing so, linguistic models provide an excellent starting point. The simplest prediction is that contextual constraints form a part of the selection mechanism itself, for example where a discourse-new status should lead to the selection of an H* accent and/or unreduced pronunciation.

This kind of mechanism is similar in spirit to theories of referential form variation in which a particular linguistic form is chosen on the basis of its pragmatic appropriateness (e.g., Ariel, 1990; Brennan, 1995; Grosz et al., 1995; Gundel et al., 1993). For example, in the Givenness Hierarchy (Gundel et al., 1993), an information status of “in focus” licenses the appropriate use of unstressed pronouns. Although this theory is primarily aimed at predicting the appropriate use of lexical forms (e.g., indefinite NPs, definite NPs, or pronouns), the inclusion of unstressed pronouns acknowledges the assumption that acoustically reduced forms are selected by highly focused discourse statuses.

The core property of these theories is that they propose that acoustic forms are chosen on the basis of information-status. Although these models do not exclude the possibility that other constraints matter too, they emphasize the importance of discourse-context constraints. A separate question concerns how these are instantiated in terms of psycholinguistic processes.

Processing models of message-based accounts—The previous section describes a class of theories that can be categorized together in that they all account for acoustic variation as a function of selectional constraints. In these theoretical traditions, research seeks to identify the contexts that allow the selection of a reduced form, or a phonological rule that leads to reduction. Here we consider how these selectional constraints might be implemented in a processing model. To our knowledge, there are no existing processing models that specify the processes of acoustic reduction. We therefore propose a few possibilities, building on models of other language production processes.

Most current models of language production assume that language production begins with a nonlinguistic representation of the to-be-uttered message, termed the Message Level. This stage is followed by linguistic formulation, which includes word selection, syntactic assignment, and the construction of the phonological form (Garrett, 1975; Levelt 1989; see Bock & Levelt, 1999 and Ferreira & Engelhardt, 2006 for a review). Although no model fully implements the role of selectional constraints, Kahn & Arnold (2012) proposed that the existence of conditioning contexts seem to require a model in which the critical contextual

¹We thank an anonymous reviewer for this point.

features are linked to a triggering mechanism that selects the appropriate form or rule. They therefore termed this class of models **trigger models**. This proposal is built on proposals that the discourse context licenses certain forms (e.g., Gundel et al., 1993). However, it goes beyond mere licensing, based on evidence that the discourse context does more than indicate which forms are allowed – it leads to reliable preferences (Arnold, 1998; 2010)². The critical property of trigger models is that they seem to require an explicit representation of the input conditions for form selection. What might this representation look like?

Input conditions: how is information status represented?—The defining characteristic of a trigger model is the representation of the information-status conditions under which particular forms are chosen. As reviewed in the previous section, scholars have described a rich set of information contrasts that correlate with acoustic form. An obvious way to implement these would be through an explicit representation of discourse status, focus structure, and/or predictability. If common ground is a critical feature of discourse status, listener knowledge could be included in this explicit representation as well.

Within current production models, the most natural locus for the representation of information status would be at the message level, or possibly at a pre-message conceptual level of representation. The precise format of this representation is an open question. For example, some models have proposed that discourse and situational information is represented in nonlinguistic mental models, sometimes termed situation models (Bower & Morrow, 1990; Bransford, Barclay, & Franks, 1972; Morrow et al., 1987, 1989; Johnson-Laird, 1983; van Dijk & Kintsch, 1983). Centering theory suggests that that discourse entities are represented as a list of entities (the “forward-looking centers”) that are rank-ordered by a syntactic hierarchy (Brennan, Friedman, & Pollard, 1987; Grosz, et al., 1995). Discourse Representation Theory posits a symbolic representation of discourse entities that is linked to the text by means of construction rules (Kamp & Reyle, 1993, see Gordon & Hendrick, 1998 for a DRT account of reference). While these proposals have substantive differences, they critically require an independent representation of discourse entities from linguistic form. Thus, a conceptual or discourse representation of referents would be necessary to represent a process by which discourse focus (or other contextual criteria) is used to select acoustically reduced forms.

A very similar idea has been developed for models of pronoun selection. For example, Schmitt, Meyer, & Levelt (1999) propose a production model in which discourse accessibility is represented by means of an “in focus” node, which serves to select for the use of a pronoun. A simple on/off representation of discourse focus is incompatible with evidence that both pronoun production and acoustic reduction are influenced by fine-grained distinctions in the discourse context, drawing on the syntactic and semantic roles of previous references, discourse expectations, and other constraints (Ariel, 2001; Arnold, 1998, 2001, 2008a, 2010; Kehler, Kertz, Rohde, & Elman, 2008; Stevenson, Crawley, & Kleinman, 1994). Nevertheless, it represents a tractable method for representing information status explicitly, and could either be expanded to include additional constraints, or could be

²Models that license possible forms (as opposed to selecting the preferred form) might be instantiated in a similar processing model, through the selection of a category of referential forms, as opposed to a single form.

considered to be the output of these constraints. Arnold & Griffin (2007) similarly adopt the view that discourse accessibility triggers pronoun selection, but assume that accessibility is gradient and subject to processing constraints like competition (see also Fukumura & van Gompel, 2010).

The explicit representation of discourse status extends to a very different kind of formalism, proposed by van Rij, van Rijn, & Hendriks (2011a, 2011b). They use the architecture of ACT-R (Anderson, 2007) in which entities from the discourse are represented as “chunks”, which are represented in declarative memory. Subject bias effects are modeled by assuming that the subject of the previous sentence is a source of activation that spreads to the chunk for that entity in memory (contingent on sufficient working memory resources), leading the model to select it as the current discourse topic. Reference forms are selected on the basis of two constraints: 1) a general preference for pronouns, for reasons of efficiency, and 2) a mechanism by which the production model checks the likely comprehension of the pronoun; when the referent is not the discourse topic, a pronoun would lead to the incorrect interpretation and is thus rejected. This model thus combines an explicit representation of discourse topic and an audience design mechanism.

Output: selection of acoustic forms—Whatever the representation of information status, there are numerous possible processes of selecting the output, including both the categorical selection of forms or processes, and the specification of a gradient process.

A natural extension of the pronoun models described here would be to instantiate accenting choices in terms of a categorical choice, possibly at the level of the intonational phrase, rather than the word. This would fit well with accounts that link acoustic meaning to categorical choices between accent types (e.g., Pierrehumbert & Hirschberg, 1990). Under this view, specific acoustic parameters like duration, pitch, and intensity, are generated on the basis of a mediating representation of prosodic structure. A second possibility would be to link information status directly with the acoustic parameters that underlie accenting, duration, pitch, and intensity (Breen et al., 2010; Lieberman, 1960). Under this view, the degree of information status (e.g., level of accessibility) might select different levels of duration or pitch movement.

In addition to accenting variation, acoustic form is influenced by alternations between full and reduced pronunciations such as t/d deletion or reduction, schwa deletion, and assimilation (Bürki et al., 2010; Ranbom & Connine, 2007; Raymond, Dautricourt, & Hume, 1995). These alternations also might be accounted for by either categorical selection or a phonological process of reduction. Bürki and colleagues have argued that some pronunciation variants are represented explicitly in the lexicon, but some are not. For example, Bürki et al. (2010, 2011) examined the pronunciation of schwa words in French and English, which can be pronounced either with the schwa present (mack-e-rel) or deleted (mack’rel). They argued that each pronunciation is separately represented in the lexicon, in contrast with models in which one pronunciation is represented and the other is derived via production rules (e.g., Côté & Morison, 2007; Tranel, 1981; F. Dell, 1985). This categorical distinction between variants would be naturally modeled with a trigger mechanism, since the speaker would need some means of selecting between them. On the other hand, Bürki &

Gaskell (2012) argue that pre-stress schwa words in English (e.g., *salami*) have only a single representation. While Bürki and colleagues do not discuss information status constraints on schwa alternations, Jurafsky, Bell, Gregory, and Raymond (2001) demonstrate that a related alternation, t/d deletion in English, is sensitive to many of the same constraints as durational reduction, which suggests that segmental deletion and other reduction processes are related.

In sum, trigger models for acoustic form provide a natural mechanism for modeling the relationship between information structure and acoustic reduction. We have argued that a trigger model is characterized by an explicit representation of discourse and/or focus status, which is used to select a particular output form. This kind of model builds naturally upon many linguistic proposals about prosody (e.g., Pierrhumbert & Hirschberg, 1990; Rooth, 1992). We have reviewed a range of input representations and output processes consistent with a trigger model; further work is needed to identify which of these accounts best for the data.

We expect that this sort of mechanism is responsible for at least some acoustic variation in language use. Language users have a strong sense that there is an “appropriate” way to say things. Nevertheless, before we can conclude that a trigger mechanism is the best model, we must consider how facilitation processes can also account for acoustic variation.

The processing facilitation approach to explaining acoustic prominence

A second approach to understanding acoustic form variation stems from the observation that acoustic prominence is systematically related to the ease of producing a word or utterance. Under this view, acoustic variation is a side effect of facilitation within the production system and not the result of pragmatic selection.

This view is consistent with the fact that the information statuses that lead to reduction are precisely those that the speaker has experience with: given, highly focused, accessible information. That is, when information is given, the speaker has heard or thought of it before, which means it is primed. There is extensive evidence that primed information is processed more quickly in dialogue (see Pickering & Garrod, 2004, for discussion). This faster processing is likely to lead to faster articulation.

Empirical support for this idea comes from extensive evidence that facilitation results in acoustic reduction (but also see Damian, 2003; Ferreira, 2007; Ferreira & Swets, 2002). This reduction is not the result of a change in rate of speech, but is specific to the information that is facilitated (e.g., Bell et al., 2009). For example, Balota et al. (1989) found that subjects were faster to pronounce printed words (e.g. *dog*) in the context of a semantically-related prime (*cat*) than control conditions. They also found that pronunciation of a target (e.g. *piano*) was shortest when an ambiguous prime (e.g., *organ*) was put in a context that evoked the meaning that was related to the target word, compared to a context that evoked an alternate meaning.

While production facilitation leads to shorter word durations, difficulty with processing is associated with longer durations. In a corpus analysis, Bell et al. (2003) found that words were longer when they adjoined disfluent elements like *um*, *uh* than when they did not.

Similarly, Clark and Fox Tree (2002) report that disfluency is associated with prolonged (i.e. longer) pronunciations (see also Arnold, Hudson Kam, & Tanenhaus, 2007). These findings support a model in which there is a relationship between word duration and the speed or facilitation of production processes.

The facilitation approach is also consistent with evidence that acoustic reduction is primarily driven by the speaker's own experience, rather than available information about the addressee's experience. For example, Bard and colleagues (Bard et al., 2000; Bard & Aylett, 2004) have found that speakers reduce repeated words, regardless of whether the addressee was present for the first mention of the word (see also Kahn & Arnold, under review). This finding supports their argument that the primary mechanism of acoustic reduction cannot be the consideration of the addressee's needs. While several alternate mechanisms are possible, a speaker-internal facilitation account is a likely one.

The facilitation of utterance planning has also been shown to affect word duration in experimental studies. For example, Gillespie (2011) analyzed spoken sentences where a noun was followed by a prepositional phrase that was either semantically integrated with the head noun, or not: *The sweater with the tiny holes ... vs. The sweater with the clean skirt ...*. They found that word durations were shorter in the integrated condition, especially over the word *the* following the preposition. Under the assumption that semantic integration facilitates planning, these results support the idea that planning leads to durational reduction.

Further evidence for planning effects on word duration comes from Christodoulou (2012). He examined how word duration varies according to whether the speaker has already begun to plan the following word. Speakers named two pictures without pausing, e.g. *toaster giraffe*. When the speaker looked at the second picture before beginning word 1, word durations were shorter. This effect interacted with the frequency of word 2, such that the effect of word 2 frequency was only observed when the speaker had fixated object 2 before speaking word 1. When speakers didn't begin planning word 2 until after initiating word 1, durations were long overall.

These planning effects suggest that durational reduction is associated with the pre-activation of words and phrases. This idea also accounts for findings that words are reduced when their referent is predictable within a discourse context (Kahn & Arnold, 2012; Lam & Watson, 2010; Watson et al., 2008). When speakers know what they are going to say, they can devote more resources to utterance formulation, which speeds activation of necessary representations, and allows for faster articulation.

Although most of the evidence about facilitation concerns duration variation, there is modest support for the idea that other correlates of acoustic prominence are also modulated by production facilitation. For example, Christodoulou (2009) had participants give instructions to a partner to click on colored pictures of either simple objects (e.g., *Click on the blue house*) or complex novel objects (e.g., *Click on the blue abstract picture that looks like a sunset over a lake*). In an analysis of only the preamble (*Click on the blue*), he found that word duration was longer preceding the hard-to-name picture. Moreover, average pitch on the color word was higher for complex objects (see Arnold et al., 2007, for similar results).

Likewise, increased pitch is characteristic of the speech patterns associated with disfluency, i.e. “thinking prosody” (Arnold & Tanenhaus, 2012). In particular, words that are repaired errors have a higher fundamental frequency than the error (Shriberg, Bear, & Downing, 1992). Nevertheless, further work is needed to understand how facilitation affects pitch and intensity, as well as duration variation.

Processing models of facilitation-based accounts—Researchers have proposed several mechanisms for facilitation-based reduction, although further research is needed to develop these more. We discuss these here under three categories: 1) Residual activation or routinization; 2) Anticipatory activation; 3) Fluency maintenance. These proposals are couched within assumptions from current models of language production, which agree that sentence production involves the activation of representations at different levels. Speakers start with a semantic representation of the intended message, from which they activate syntactic, morphological, phonological, and articulatory representations (Bock, 1986; Bock & Levelt, 1999; Dell, 1986; Ferreira & Englehart, 2006; Garrett, 1975 1980; Levelt, 1989; Levelt, Roelofs & Meyer, 1999).

One proposal is that activation at conceptual and lexical levels leads to faster activation of phonological and articulatory representations, which translates into faster pronunciation of a word. For example, Balota et al. (1989) reported that words are pronounced more quickly when they are primed. A related idea (Bybee, 2001) is that frequent words are shorter because the articulatory processes are routinized (but see Gahl, 2008). Kahn & Arnold (2012) have found that in addition to linguistic givenness, nonlinguistic-conceptual givenness can contribute to reduction, suggesting that reduction may be the result of facilitation from multiple levels of representation in the language production system.

Another proposal draws on the fact that given and accessible information does not just have the property of having already been used – it also tends to be information that is repeated later in the discourse. That is, it is relatively predictable (Arnold, 1998; 2010; Grosz et al., 1995; Lam & Watson, 2010; Prince, 1981; Watson et al., 2008). If predictability of a particular word is high enough, the speaker might start planning it earlier. Even partial predictability could lead speakers to maintain the activation of previously-encountered items, based on the likelihood that the item will be re-mentioned.

A final proposal is that speakers have a coordination mechanism that allows them to slow articulation of a word when lexical access of a subsequent word is delayed (Bell et al., 2009, but see Goldrick, Baker, Murphy, & Baese-Berk, 2011). This mechanism would presumably serve the social goal of maintaining fluency, despite the information processing demands of utterance production. This mechanism accounts for why facilitation effects appear on adjoining words, and not just the facilitated word itself.

The notion that articulation time might be linked to planning and lexical activation is consistent with two-stage models of word production. The first stage requires accessing the lexical entry for the word while the second stage requires assembling the word’s phonological form (e.g. Dell, 1986; Garrett, 1988; Levelt 1989). Data from the word production literature suggests that the process of assembling the phonological form of a

word occurs serially with initial segments of a word activated first (e.g. Meyer, 1991; Sevald & Dell, 1994). For example, Sevald and Dell (1994) found that repetitions of pairs of words that shared initial phonemes took longer to articulate than pairs of words that shared final phonemes (see O'Seaghda & Marin, 2000 for similar results). Sevald and Dell (1994) proposed a model in which phonemes are activated sequentially and send feedback to higher-level lexical nodes. This activation of lexical nodes creates competition between lexical competitors at the beginning of articulation, which leads to slow-downs in productions. Although there is also feedback to higher lexical levels when word pairs share final phonemes, this competition occurs only after most of the initial processing is complete. In the context of theories of lexical production, extending word duration for new or low probability words might provide more time for the process of phonological assembly to take place. Critically, given the sequential nature of phonological production, speakers would benefit from lengthening at the point at which the word is actually being produced. Conversely, if the processes involved in phonological assembly have been primed because of repetition, the process could run more quickly and result in reduction.

ONE SYSTEM OR TWO?

The preceding section proposes that there are two classes of explanations for variation in acoustic prominence as a result of information status: one in which discourse status is explicitly represented (the trigger account), and another in which the discourse context has concomitant facilitation effects on linguistic form (the facilitation account). How should we think of these accounts? Do they simply reflect two different kinds of effects, or are they a part of the same system?

On the separate systems view, trigger and facilitation processes are unrelated, except for the fact that they affect the same outcome measures. One version of this view stems from a theoretical interest in the message-based system. The competence/performance distinction emphasizes the fact that language users have knowledge of the grammatical or felicitous way to say things, even if performance factors also affect their ability to use this knowledge. On this view, facilitation-based phenomena should be distinguished from linguistic knowledge, simply because they are not the focus of inquiry. Another version of this view is that both facilitation and trigger mechanisms influence the output, but that they should be viewed as separate systems. This is the perspective taken by Ferreira (2007), who examines this question with respect to intonational breaks and pauses. When we consider the evidence on acoustic prominence, one argument in favor of separate systems is that processing effects have primarily been reported for timing and duration measures, whereas meaningful constraints on prominence tend to be characterized in terms of pitch accent. Nevertheless, accent is at least partially encoded by durational variation (Ladd, 1996), blurring the line between these effects. Moreover, few studies have explicitly examined the effects of processing constraints on pitch variation, which means that further investigation is needed to understand this relationship.

A second view, which we argue for, focuses on the fact that the mechanisms by which the discourse context affects acoustic reduction are not yet well understood. Therefore, it is entirely possible that known discourse effects on prosody are mediated by processing

facilitation. That is, perhaps some discourse effects result from the processing facilitation that is concomitant with given or predictable discourse status.. This means that an understanding of facilitation constraints on prominence is a necessary component of research on acoustic prominence, even if the goal of the research is only to understand the semantic and pragmatic constraints. If processing factors are not controlled for, a facilitation mechanism cannot be ruled out.

Our reasoning is based on two facts: First, information status categories pattern with facilitation, since given/accessible information is systematically easier to produce than new information. Second, the predicted effects of pragmatic mechanisms (“choose reduction for given information”) and facilitation mechanisms (“reduction happens when production is easier”) lead to overlapping acoustic prominence profiles. Another way of saying this is that information status is correlated with processing facilitation.

The correlation between information status and processing facilitation

From a processing perspective, it makes sense that on average, given or predictable information should be easier to produce than new or unpredictable information. Simply put, it is easier to talk about information that is already active. Given information has already been evoked in some way (Prince, 1981), which means that it has been recently activated conceptually, and in most cases has been mentioned linguistically as well. Psycholinguistic research has shown that for many phenomena, recently produced information is easier and faster to produce again (e.g., Bock, 1986; Levelt, Roelofs & Meyer, 1999). The relationship between information status and production facilitation is supported by research on disfluency, which can be considered an index of production difficulty. If given information is easier to mention, given references should be more fluent than new references. and indeed they are (Arnold & Tanenhaus, 2012). Thus, information status of a word is likely to correlate with the ease of producing that word. Speakers can also begin formulating predictable information earlier, which should also facilitate production. This correlation is not perfect, in that other lexical and situational properties modulate difficulty. Moreover, different kinds of givenness correlate with processing facilitation to different degrees (Kahn & Arnold, 2012). Yet this correlation is likely to be robust enough that it merits investigation.

Processing facilitation may even be related to audience design effects. Many theories suggest that speakers make prosodic choices on the basis of audience design – that is, with respect to their assumptions about the addressee’s knowledge or attention (Pierrehumbert & Hirschberg, 1990; see Gundel et al., 1993, for a similar argument about lexical variation). On this view, the speaker chooses acoustically prominent forms because either the information is new to the addressee, or because the addressee is expected to have trouble retrieving the word or meaning. However, explicit tests of this assumption have found mixed results (Arnold, et al., 2012; Bard et al., 2000; Bard & Aylett, 2004; Kahn & Arnold, under review; Galati & Brennan, 2010; Rosa et al., this volume). Arnold et al. (2012) suggest that in some cases, the available evidence about the addressee’s knowledge or attention may affect the speaker’s attention, and thus facilitate or inhibit production planning (see also Kahn & Arnold, under review, this volume). On this view, the addressee is one source of

relevant information. The addressee's behavior can provide evidence about what the addressee knows, which can affect processing facilitation. That is, addressee knowledge and behavior can also direct the speaker's production processes.

In sum, evidence suggests an inter-relationship between facilitation and both information status and audience design. This relationship has several consequences. Most broadly, it poses a data interpretation problem, especially for researchers who aim to provide a mechanistic account of acoustic variation in language production. We know that speakers tend to use reduced pronunciations for given, focused, or predictable information. But as we have shown, there are two types of production mechanism that could yield these results, either one based on contextually-based selection of acoustic forms (a trigger mechanism), or one based on processing difficulty or facilitation (a facilitation mechanism). This calls for research to tease apart these accounts, as outlined in the next section.

Future directions: Synthesizing facilitation and message-based accounts

Given that processing facilitation and information status are correlated, we propose that the field do two things. First, we need theoretical accounts of language production that incorporate fine-grained characterizations of information status within a mechanistic model of production processes. This is a general problem that exists for both message-based and facilitation-based approaches.,

Second, we need empirical studies that identify the relationship between pragmatic constraints and the concomitant facilitation mechanisms, and the effects of each on acoustic prominence. Researchers cannot observe a single phenomenon – say, the tendency for speakers to reduce repeated words – and assume that it reflects either pragmatic or processing constraints alone. We expect that a full model of acoustic reduction may involve elements of both models, but until we examine this question directly, it is impossible to deconvolve message-based and facilitation-based explanations.

Here we outline a few research directions that would be fruitful for understanding whether and how processing facilitation mediates observed pragmatic effects on acoustic reduction.

Discourse status effects on Planning—More research is needed to understand how discourse status affects word and utterance planning, and the effects of planning on acoustic prominence. For example, the lemmas and word forms associated with given information are likely activated, leading to faster re-activation on re-mention. How does this activation relate to the time-course of planning?

One approach to this problem is to use current methods in language production (e.g., eye-tracking or priming techniques) to map out the time-course of utterance planning in different discourse situations. We have proposed that processing facilitation may underlie some discourse effects; one specific version of this hypothesis is that this facilitation stems from conceptual or grammatical planning. Given and accessible discourse statuses are highly likely to affect utterance planning, since they are linked to the predictability of later re-mention (Arnold, 1998, 2001, 2010; Givon, 1983; Grosz, et al., 1995; Watson, et al., 2008).

Eyetracking can be used to identify the time-course of planning words that refer to given and new information, and associating planning time with acoustic reduction.

Another approach to examining discourse effects on planning is to individually examine each property associated with information status. There are multiple proposals for how information status is structured (for some examples, see Chafe, 1994; Prince, 1981; Rooth, 1992; Vallduvi, 1993), reflecting the multidimensional nature of these categories. For example, given information tends to generally (but not always) have these characteristics: being conceptually active and attended in the mind of the speaker, being conceptually active and attended in the mind of the addressee(s), having been recently mentioned, and thus having been either articulated or heard by the speaker and addressee, having played a topical role in the preceding discourse, and having a high likelihood of continued mention in the discourse. It may also be visually salient, and related to the task goals. Any or all of these characteristics may independently contribute to processing facilitation. By decomposing givenness into token-specific characteristics, researchers can get a handle on ways in which processing facilitation may be independent of a general category of givenness.

This approach underlies Kahn and Arnold's (2012) proposal that facilitation at any processing level within the production system should contribute to acoustic reduction, such that facilitation at multiple levels leads to more reduction than facilitation at just one level. They tested this idea by examining how speakers refer to objects that are given and highly accessible, where that givenness is achieved either linguistically or non-linguistically. Since linguistic givenness should result in activation for both conceptual representations and linguistic ones (e.g., lexical and phonological), linguistic givenness should result in greater acoustic reduction than conceptual givenness. The results of two experiments confirmed this prediction. On the other hand, subsequent work (Kahn & Arnold, under review) demonstrated only modest support for the hypothesis that articulatory experience contributes to reduction.

In principle, both trigger and facilitation models can account for the fact that different kinds of givenness affect acoustic variation to different degrees. However, as argued by Kahn & Arnold (2012), facilitation offers a simpler approach, whereas a trigger model would require the explicit representation of visual vs. linguistic givenness, as well as any other distinction that matters (see, e.g., Prince, 1981). Further work is needed to map out the contribution of different characteristics associated with discourse givenness or accessibility.

Examine different measures of acoustic reduction—Another route to understanding the mechanisms of acoustic reduction is to observe that different acoustic measures do not always pattern together. One possibility is that the primary predictor of reduction is the choice in accenting, such that unaccented forms are acoustically reduced. This would be most consistent with a trigger mechanism in which an accent category (e.g., H*) is selected, and this in turn selects the acoustic instantiation of the accent in terms of duration, pitch, and intensity. This would not predict that different discourse situations would affect each acoustic parameter differently. In contrast to this, Lam & Watson (2010) have shown that discourse repetition and predictability independently affect the acoustic signal in different ways (see also Watson, 2010). While both factors affect a word's intensity

and duration, repetition accounts for most of the variance in duration, while predictability accounts for most of the variance in intensity. More research in this line is needed to distinguish between a trigger mechanism that directly controls duration and pitch, and a facilitation mechanism in which each acoustic parameter is differentially affected.

Focus effects and facilitation—Another promising avenue is to identify phenomena that seem to demand trigger but not facilitation mechanisms. One of the biggest challenges for facilitation accounts is to explain contrastive accenting on given words (see also Wagner & Klassen, under review). For example, in *The cat and dog were running, and then the cat jumped*, the second *cat* is likely to be even more prominent than the first one. In fact, pilot data from the first author's lab suggests that this is so, and even that a contrastive noun is more prominent than a target that is discourse-new. This acoustic prominence is predicted by a mechanism in which contrast triggers acoustic prominence (Baumann, Grice, & Steindamm, 2006; Krahmer & Swerts, 2001). The word *cat* is a member of the contrast set including the cat and dog, and picking out just the cat implies a contrast with the dog.

A critical question is whether facilitation contributes at all to this prominence. At the lexical level, prior mention of the word *cat* should facilitate reuse of the word in the next clause, which predicts reduction, not prominence. Yet other levels of processing may not be facilitated. At the referential level, *the cat* is introduced as a member of a compound NP, casting the cat and dog as a single entity. Subsequent reference to this entity is predictable, but mention of just one part may be even more difficult, because it requires conceptual restructuring. To test this question, production experiments could examine the degree of acoustic prominence when other information in the context encourages the referential integration of the two elements, vs. when it does not (cf. Brown-Schmidt, Byron, & Tanenhaus, 2005; Gillespie, 2012). Yet if facilitation has only a partial influence (if any), this line of research could place a limiting constraint on the role of facilitation.

Examine non-discourse constraints on reduction—To test the hypothesis that pragmatic effects on acoustic reduction are mediated by facilitation mechanisms, it is also important to examine non-discourse manipulations that affect acoustic reduction, and how they interact with discourse effects. Acoustic reduction is likely to occur when speakers are paying attention to the task/referent (cf. Rosa & Arnold, 2011), and when they pre-plan their utterances. Even word frequency effects, which are well established (e.g., Bell et al., 2009), cannot be accounted for by discourse constraints per se.

The critical question, outlined above, is whether these are part of the same system or different systems. We have suggested that both discourse and frequency effects may stem from processing facilitation mechanisms, predicting potential interactions. If both givenness and frequency effects are partly driven by facilitation, then they might be smaller for frequent words, since frequent words are already relatively easy to say. In order to test this hypothesis, experimental manipulations or corpus analysis needs to tightly control the categorization of givenness. Recently mentioned words are likely to have a much stronger effect on reduction than less-recently mentioned words (cf. Arnold, 1998). Therefore, a binary categorization that contrasts brand-new words with those mentioned somewhere previously in the discourse (e.g., Bell et al., 2009) is likely to result in a noisy and weak

measure of repetition effects, reducing the possibility of identifying any additional interactions.

Which words are reduced?—Related to the planning questions discussed above, a final issue that needs exploration is the fact that message-based and facilitation-based accounts make different predictions about the location of reduction. Message-based accounts suggest that accents are assigned to specific words or constituents. For example, *the SESAME bagel* focuses the word “sesame” in comparison with the surrounding words. Thus, a relatively rapid pronunciation could sound prominent in the context of a fast speech rate overall, whereas a longer one might sound reduced in the context of very slow speech. Under this view, prominence is a relative phenomenon. As such, it is often measured perceptually, as listeners can identify the most prominent-sounding segment (e.g., Terken & Hirschberg, 1994).

By contrast, processing facilitation can result in acoustic reduction over a range of words, depending on the time course of planning (Bell et al., 2003; Bell et al., 2009; Kahn & Arnold, under review). For example, Gillespie (2011) found that semantic integration led to reduced word duration over a region spanning the first three words of the prepositional phrase; e.g., in the phrase “the sweater with the tiny holes”, the words *with*, *the*, and *tiny* were reduced. These findings emerge in acoustic analyses of duration, which focus on absolute rather than relative variation. Thus, the location of facilitation and message-based effects overlap, but may have a different signature.

CONCLUSIONS

We have shown that acoustic prominence is conditioned by multiple competence and facilitation constraints. Although the acoustic properties of each type may differ somewhat, there is also considerable overlap. Therefore, the acoustic signal is at least partially ambiguous: it can signal processing ease/difficulty, and it can also signal discourse or grammatical information (Clifton et al., 2006, Ferreira, 2007).

The systematic relationship between performance and competence constraints muddies the waters for researchers interested in figuring out the structure of prosody. It also raises a potential problem for the listener. We know that language comprehension is guided by both types of constraint. On the one hand, listeners rapidly use message-based acoustic prominence, for example when it signals discourse status (Arnold, 2008b), or a contrast set (Eberhard et al., 1995). On the other hand, pitch excursions may be one of the perceptual indicators of disfluency, which creates a rapid on-line bias toward a difficult-to-name object (Arnold, Tanenhaus, Altmann & Fagnano, 2004; Arnold, Hudson-Kam, & Tanenhaus, 2007, Arnold & Tanenhaus, 2012). If prominence reflects processing difficulty that is unrelated to information status, do listeners misinterpret the speaker’s intended meaning? This problem also extends to acquisition: how do children who are learning the prosody of their language disentangle grammatical information about prosody from performance constraints?

Because of this correlation, we have highlighted the need for the field to directly examine the relative contribution of message-based and facilitation-based mechanisms. We propose

that the only way to do this effectively is to consider actual processing mechanisms. We suggested that message-based selection falls into a class of model that we have termed Trigger models (following Kahn & Arnold, 2012). These contrast with a class of model that we have termed Facilitation models, in which acoustic variation does not depend directly on an explicit representation of the conditioning contexts.

We have also proposed that message-based and facilitation-based mechanisms should be considered as part of the same system, and not two separate systems. By this we mean that pragmatic effects may even be partially mediated by a facilitation mechanism. That is, facilitation may contribute to the forms associated with particular information statuses. In that sense, there may only be one system, encompassing both the effects of the discourse and the facilitation mechanisms that underlie them. We are not simply arguing that performance and competence constraints on prominence need to be distinguished (cf. Ferreira, 2007). Nor are we arguing that grammar codifies performance constraints (e.g., Walters, 2007), or that certain grammatical constraints are learned more easily (Moreton, 2008).

Rather, we are making a strong claim that in some cases the acoustic consequences of competence constraints are mediated by performance – i.e. facilitation – mechanisms. This means that the opposite view of separate mechanisms needs to be supported empirically, not assumed. Likewise, this processing-mediated view may be falsified if a trigger mechanism is shown to underlie all information status effects. Even if specific support emerges for a trigger mechanism, we predict that at least some discourse effects result from processing facilitation.

We have outlined some specific research directions that would help identify the underlying mechanism. Some fruitful directions include a focus on the relationship between discourse constraints and production planning, an examination of different types of acoustic reduction, an examination of non-discourse constraints on acoustic variation, and a consideration of the parts of the utterance affected by both discourse and facilitation manipulations.

Our proposal joins existing work in both psycholinguistic and grammatical systems by assuming that the goal of language research is to understand the cognitive systems (grammatical and processing) that underlie language use. We argue that in order to understand acoustic prominence as a part of successful communication, we need to understand idealized prosodic categories in combination with, and in relation to, the psychological implementation of acoustic reduction.

Acknowledgments

This research was partially supported by NSF grant BCS-0745627 to J. Arnold. D. Watson is supported by NIH grant R01 DC008774 and the James S. McDonnell foundation.

References

Anderson, JR. How can the human mind occur in the physical universe?. New York: Oxford University Press, USA; 2007.

- Ariel, Mira. Accessibility theory: An overview. In: Sanders, T.; Schilperoord, J.; Spooren, W., editors. Text representation: Linguistic and psycholinguistic aspects. Amsterdam: Benjamins; 2001. p. 29-87.
- Ariel, M. Accessing noun-phrase antecedents. London: Routledge; 1990.
- Arnold, JE. Unpublished doctoral dissertation. Stanford University; 1998. Reference form and discourse patterns.
- Arnold JE. The Effect of Thematic Roles on Pronoun Use and Frequency of Reference Continuation. *Discourse Processes*. 2001; 31(2):137–162.10.1207/S15326950DP3102_02
- Arnold JE. Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*. 2008a; 23(4):495–527.10.1080/01690960801920099
- Arnold JE. THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition*. 2008b; 108(1):69–99.10.1016/j.cognition.2008.01.001 [PubMed: 18358460]
- Arnold JE. How speakers refer: The role of accessibility. *Language and Linguistics Compass*. 2010; 4(4):187–203.10.1111/j.1749-818X.2010.00193.x
- Arnold, JE.; Tanenhaus, MK. Disfluency isn't just um and uh: The role of prosody in the comprehension of disfluency. In: Gibson, E.; Perlmutter, N., editors. The processing and acquisition of reference. Cambridge, MA: MIT Press; 2012.
- Arnold JE, Griffin ZM. The effect of additional characters on choice of referring expression: Everyone counts. *Journal of memory and language*. 2007; 56(4):521–536.10.1016/j.jml.2006.09.007 [PubMed: 18449327]
- Arnold JE, Kam CLH, Tanenhaus MK. If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology Learning, Memory, and Cognition*. 2007; 33(5):914–30.10.1037/0278-7393.33.5.914
- Arnold JE, Tanenhaus MK, Altmann RJ, Fagnano M. The old and thee, uh, new: disfluency and reference resolution. *Psychological Science*. 2004; 15(9):578–582.10.1111/j.0956-7976.2004.00723.x [PubMed: 15327627]
- Arnold JE, Kahn JM, Pancani GC. Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin and Review*. 2012;10.3758/s13423-012-0233-y
- Aylett M, Turk A. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*. 2004; 47:31–56. [PubMed: 15298329]
- Balota DA, Boland J, Shields L. Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*. 1989; 28:14–36.
- Bard, EG.; Aylett, MP. Referential form, word duration, and modeling the listener in spoken dialogue. In: Trueswell, JC.; Tanenhaus, MK., editors. Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions. Cambridge, MA: MIT Press; 2004. p. 173-191.
- Bard EG, Anderson AH, Sotillo C, Aylett M, Doherty-Sneddon G, Newlands A. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*. 2000; 42(1): 1–22.10.1006/jmla.1999.2667
- Baumann, S.; Hadelich, K. Accent type and givenness: an experiment with auditory and visual priming. *Proceedings of the 15th ICPHS; Barcelona*. 2003. p. 1811-1814.
- Baumann S, Grice M. The intonation of accessibility. *Journal of Pragmatics*. 2006; 38(10):1636–1657.10.1016/j.pragma.2005.03.017
- Baumann, S.; Grice, M.; Steindamm, S. Prosodic Marking of Focus Domains - Categorical or Gradient?. *Proceedings Speech Prosody; 2006; Dresden, Germany*. 2006. p. 301-304.
- Bell A, Brenier J, Gregory M, Girand C, Jurafsky D. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*. 2009; 60(1):92–111.
- Bell A, Jurafsky D, Fosler-Lussier E, Girand C, Gregory M, Gildea D. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*. 2003; 113(2):1001–1024. [PubMed: 12597194]
- Bock, JK.; Levelt, WJM. *Language production: Grammatical encoding*. MA, editor. 1994.

- Bock JK, Mazella JR. Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*. 1983; 11:64–76. [PubMed: 6855561]
- Bock JK. Meaning, Sound, and Syntax: Lexical Priming in Sentence Production *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1986; 12:575–586.
- Bower GH, Morrow DG. Mental models in narrative comprehension. *Science*. 1990; 247(4938):44–48. [PubMed: 2403694]
- Bransford JD, Barclay JR, Franks JJ. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*. 1972; 3(2):193–209.
- Breen M, Fedorenko E, Wagner M, Gibson E. Acoustic correlates of information structure. *Language and Cognitive Processes*. 2010; 25(7):1044–1098.
- Brennan SE. Centering attention in discourse. *Language and Cognitive Processes*. 1995; 10:137–167.
- Brennan SE, Hanna JE. Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science*. 2009; 1(2):274–291.10.1111/j.1756-8765.2009.01019.x [PubMed: 25164933]
- Brennan, SE.; Friedman, MW.; Pollard, CJ. Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics. Cambridge, MA: Association for Computational Linguistics; 1987. A centering approach to pronouns; p. 155-162.
- Brown, G. Prosodic structures and the Given/New distinction. In: Ladd, DR.; Cutler, A., editors. *Prosody: Models and measurements*. Berlin: Springer; 1983. p. 67-77.
- Brown-Schmidt S, Byron DK, Tanenhaus M. Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*. 2005; 53:292–313.
- Bürki A, Gaskell MG. Lexical representation of schwa words: two mackerels, but only one salami. *Journal of experimental psychology. Learning, memory, and cognition*. 2012; 38(3):617–31.10.1037/a0026167
- Bürki A, Alario FX, Frauenfelder UH. Lexical representation of phonological variants: Evidence from pseudohomophone effects in different regiolects. *Journal of Memory and Language*. 2011; 64(4): 424–442.10.1016/j.jml.2011.01.002
- Bürki, A.; Ernestus, M.; Frauenfelder, UH. *Journal of Memory and Language*. Vol. 62. Elsevier Inc; 2010. Is there only one “fenêtre” in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words; p. 421-437.
- Bybee, J. *Cambridge Studies in Linguistics*. Vol. 94. Cambridge: Cambridge University Press; 2001. Phonology and language use.
- Chafe, W. Cognitive constraints on information flow. In: Tomlin, R., editor. *Coherence and grounding in discourse*. Amsterdam: John Benjamins; 1987. p. 21-51.
- Chafe, WL. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press; 1994.
- Chomsky, N. *Aspects of the theory of syntax*. The MIT Press; Cambridge, Ma: 1965.
- Christodoulou, A. Unpublished Master’s thesis. UNC Chapel Hill; Chapel Hill, NC: 2009. Thinking prosody: How speakers indicate production difficulty through prosody.
- Christodoulou, A. Unpublished doctoral dissertation. University of North Carolina; Chapel Hill: 2012. Variation in word duration and planning.
- Clark, HH.; Haviland, SE. Psychological processes in linguistic explanation. In: Cohen, D., editor. *Explaining linguistic phenomena*. Washington: Hemisphere Publication Corporation; 1974.
- Clark HH, Fox Tree JE. Interpreting pauses and ums at turn exchanges. *Discourse Processes*. 2002; 34(1):37–55.
- Clark, HH.; Marshall, CR. Definite reference and mutual knowledge. In: Joshi, AK.; Webber, B.; Sag, I., editors. *Elements of discourse understanding*. Cambridge: Cambridge University Press; 1981. p. 10-63.
- Clark HH, Wasow T. Repeating words in spontaneous speech. *Cognitive Psychology*. 1998; 37:201–242. [PubMed: 9892548]
- Clifton C Jr, Frazier L, Carlson K. Tracking the what and why of speakers’ choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin & Review*. 2006; 13:854–861. [PubMed: 17328385]

- Côté MH, Morrison GS. The nature of the schwa/zeroalternation in French clitics: Experimental and non-experimental evidence. *Journal of French Language Studies*. 2007; 17:159–186.
- Dahan D, Tanenhaus MK, Chambers CG. Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language*. 2002; 47:292–314.
- Damian MF. Articulatory duration in single word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2003; 29:416–431.
- Dell GS. A spreading activation theory of retrieval in language production. *Psychological Review*. 1986; 93:283–321. [PubMed: 3749399]
- Dell, F. *Les règleset lessons*. Paris: Hermann; 1985.
- Dell GS. A spreading-activation theory of retrieval in language production. *Psychological Review*. 1986; 93:283–321. [PubMed: 3749399]
- Eberhard K, Spivey-Knowlton M, Sedivy J, Tanenhaus M. Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*. 1995; 24:409–436. [PubMed: 8531168]
- Ferreira F. Prosody and performance in language production. *Language and Cognitive Processes*. 2007; 22:1151–1177.
- Ferreira, F.; Engelhardt, PE. Syntax and production. In: Traxler, MJ.; Gernsbacher, MA., editors. *Handbook of Psycholinguistics*. Oxford, UK: Elsevier Inc; 2006. p. 61-91.
- Ferreira F, Swets B. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*. 2002; 46:57–84.
- Fowler CA. Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*. 1988; 31(4):307–319. [PubMed: 3271893]
- Fowler C, Housum J. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*. 1987; 26(5):489–504.
- Fukumura K, van Gompel RPG, Pickering MJ. The use of visual context during the production of referring expressions. *Quarterly journal of experimental psychology (2006)*. 2010; 63(9):1700–15.10.1080/17470210903490969 [PubMed: 20127548]
- Gahl S. "Thyme" and "Time" are not homophones. Word durations in spontaneous speech. *Language*. 2008; 84(3):474–496.
- Gahl S, Garnsey SM. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*. 2004; 80(4):748–775.
- Gahl S, Yao Y, Johnson K. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language*. 2012; 66(4):789.
- Galati A, Brennan SE. Attenuating repeated information: For the speaker, or for the addressee? *Journal of Memory and Language*. 2010; 62:35–51.
- Garrett, MF. The analysis of sentence production. In: Bower, G., editor. *Psychology of learning and motivation*. Vol. 9. New York: Academic Press; 1975. p. 133-177.
- Garrett, MF. Levels of processing in sentence production. In: Butterworth, B., editor. *Language production*. Vol. 1. London: Academic Press; 1980. p. 177-220.
- Garrett, MF. Processes in language production. In: Newmeyer, F., editor. *Linguistics: The Cambridge survey: Vol 3. Language: Psychological and biological aspects*. New York: Cambridge University Press; 1988. p. 69-96.
- Gillespie, M. Unpublished doctoral dissertation. Northeastern University; Boston, MA: 2011. Agreement computation in sentence production conceptual and temporal factors.
- Givon, T. *Topic continuity in discourse: A quantitative cross-linguistic study*. Philadelphia, PA: John Benjamins; 1983.
- Gernsbacher, editor. *Handbook of Psycholinguistics*. San Diego, CA: Academic Press; p. 945-984.
- Goldrick M, Baker HR, Murphy A, Baese-Berk M. Interaction and representational integration: Evidence from speech errors. *Cognition*. 2011; 121:58–72. [PubMed: 21669409]
- Gordon PC, Grosz BJ, Gilliom LA. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*. 1993; 17:311–347.

- Gordon P, Hendrick R. The Representation Processing in Discourse of Coreference. *Cognitive Science*. 1998; 22(4):389–424.
- Gregory M, Raymond WD, Bell A, Fosler-Lussier E, Jurafsky D. The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*. 2000; 35:151–66.
- Grosz BJ, Weinstein S, Joshi AK. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*. 1995; 21(2):203–225.
- Gundel JK, Hedberg N, Zacharaski R. Cognitive status and the form of the referring expressions in discourse. *Language*. 1993; 69:274–307.
- Gussenhoven, C. A semantic analysis of the nuclear tones of English. Bloomington, Indiana: Indiana University Linguistics Club; 1983.
- Halliday, MAK. Intonation and grammar in British English. The Hague: Mouton; 1967.
- Jaeger, TF. PhD thesis. Stanford University; Stanford, CA: 2006. Redundancy and Syntactic Reduction in Spontaneous Speech.
- Jaeger TF. Redundancy and Reduction: Speakers Manage Information Density. *Cognitive Psychology*. 2010; 61 (1):23–62. [PubMed: 20434141]
- Johnson-Laird, PN. Mental models: Toward a cognitive science of language, inference and consciousness. Cambridge, MA: Harvard University Press; 1983.
- Jurafsky, D.; Bell, A.; Gregory, M.; Raymond, WD. Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In: Bybee; Joan; Hopper, Paul, editors. Frequency and the emergence of linguistic structure. Amsterdam: John Benjamins; 2001. p. 229-254.
- Kahn J, Arnold JE. A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*. 2012
- Kahn, J.; Arnold, JE. Lexical activation effects on spoken word reduction. University of North Carolina; Chapel Hill: under review, this volume
- Kamp; Hans; UweReyle. From discourse to logic. Dordrecht, The Netherlands: Kluwer; 1993.
- Kehler A, Kertz L, Rohde H, Elman J. Coherence and Coreference Revisited. *Journal of Semantics Special Issue on Processing Meaning*. 2008; 25:1–44.
- Krahmer E, Swerts M. On the alleged existence of contrastive accents. *Speech Communication*. 2001; 34:391–405.
- Ladd, R. Intonational Phonology. Cambridge: University Press; 1996.
- Lam TQ, Watson DG. Repetition is easy: why repeated referents have reduced prominence. *Memory & Cognition*. 2010; 38(8):1137–46.10.3758/MC.38.8.1137 [PubMed: 21156876]
- Lambrecht, K. Cambridge Studies in Linguistics. Vol. 71. Cambridge: Cambridge University Press; 1994. Information structure and sentence form: Topic, focus, and the mental representation of discourse referents.
- Levelt, WJM. Speaking: From intention to articulation. Cambridge, MA: MIT Press; 1989.
- Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences*. 1999; 22:1–38. [PubMed: 11301520]
- Levy, R.; Jaeger, TF. Speakers optimize information density through syntactic reduction. *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*; 2007.
- Meyer AS. The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*. 1991; 30:69–89.
- Moreton, Elliott. Analytic bias and phonological typology. *Phonology*. 2008; 25(1):83–127.
- Morrow DG, Bower GH, Greenspan SL. Accessibility and situation models in narrative comprehension. *Journal of Memory and Language*. 1987; 26:165–187.
- Morrow DG, Bower GH, Greenspan SL. Updating situation models during narrative comprehension. *Journal of Memory and Language*. 1989; 28:292–312.
- O’Seaghdha PG, Marin JW. Phonological competition and cooperation in form-related priming: Sequential and non-sequential processes in word production. *Journal of Experimental Psychology: Human Perception and Performance*. 2000; 26:57–73. [PubMed: 10696605]
- Pickering MJ, Garrod S. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*. 2004; 27:169–206. [PubMed: 15595235]

- Pierrehumbert, J.; Hirschberg, J. The meaning of intonational contours in the interpretation of discourse. In: Cohen, PR.; Morgan, J.; Pollack, ME., editors. *Intentions in communication*. Cambridge, MA: MIT Press; 1990. p. 271-311.
- Prince, E. Toward a taxonomy of given-new information. In: Cole, P., editor. *Radical Pragmatics*. NY: Academic Press; 1981. p. 223-256.
- Prince, E. The ZPG letter: subjects, definiteness, and information-status. In: Thompson, S.; Mann, W., editors. *Discourse description: diverse analyses of a fund raising text*. Philadelphia/Amsterdam: John Benjamins B.V; 1992. p. 295-325.
- Ranbom LJ, Connine CM. Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*. 2007; 57:273–298.
- Raymond WD, Dautricourt R, Hume E. Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*. 2006; 18(01):55–97.10.1017/S0954394506060042
- Rooth M. A theory of focus interpretation. *Natural Language Semantics*. 1992; 1:75–116.
- Rosa E, Finch K, Bergeson M, Arnold JE. The Effects of Addressee Attention on Prosodic Prominence. *Language and Cognitive Processes*. 2013.1080/01690965.2013.772213
- Schmitt BM, Meyer aS, Levelt WJ. Lexical access in the production of pronouns. *Cognition*. 1999; 69(3):313–35. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10193050>. [PubMed: 10193050]
- Schwarzschild R. GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics*. 1999; 7:141–177.
- Selkirk, EO. Sentence prosody: Intonation, stress and phrasing. In: Goldsmith, JA., editor. *The Handbook of Phonological Theory*. Oxford: Blackwell; 1996. p. 138-151.
- Sevald CA, Dell GS. The sequential cuing effect in speech production. *Cognition*. 1994; 53:91–127. [PubMed: 7805353]
- Shattuck-Hufnagel, Sevald CA, Dell GS. The sequential cuing effect in speech production. *Cognition*. 1994; 53:91–127. [PubMed: 7805353]
- Shriberg, EE.; Bear, J.; Dowding, J. Automatic detection and correction of repairs in human-computer dialog. In: Marcus, M., editor. *Proc. DARPA Speech and Natural Language Workshop*; Harriman, NY. 1992. p. 419-424.
- Steedman M. Information structure and the syntax-phonology interface. *Linguistic Inquiry*. 2000; 31:649–689.
- Stevenson RJ, Crawley RA, Kleinman D. Thematic roles, focus and the representation of events. *Language and Cognitive Processes*. 1994; 9:473–592.
- Terken J, Hirschberg J. Deaccentuation of words representing “given” information: Effects of persistence of grammatical function and surface position. *Language and Speech*. 1994; 37:125–145.
- Tily H, Gahl S, Arnon N, Snider N, Kothari A, Bresnan J. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*. 2009; 1(2):147–165.
- Tranel, B. *Concreteness in generative phonology: Evidence from French*. Berkeley, CA: University of California Press; 1981.
- Vallduví, E. *Information packaging: A survey*. Report prepared for Word Order, Prosody, and Information Structure. Centre for Cognitive Science and Human Communication Research Centre, University of Edinburgh; 1993.
- Van Rij, J.; van Rijn, H.; Hendriks, P. WM load influences the interpretation of referring expressions. *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*; Portland, OR. 2011a. p. 67-75.
- vanRij, J.; van Rijn, H.; Hendriks, P. Towards a cognitively plausible computational model of reference. In: van Deemter, Kees; Gatt, Albert; van Gompel, Roger; Krahmer, Emiel, editors. *Proceedings of the PRE-CogSci 2011 workshop on Production of Referring Expressions: Bridging the gap between computational, empirical & theoretical approaches*; 2011b.
- Wagner M, Klassen J. Accessibility is no Alternative to Alternatives. under review, this volume.

- Walters, MA. PhD dissertation. Massachusetts Institute of Technology; 2007. Repetition Avoidance in Human Language.
- Watson, DG. The many roads to prominence: Understanding emphasis in conversation. In: Ross, B., editor. *The Psychology of Learning and Motivation*. Vol. 52. Elsevier; 2010. p. 163-183.
- Watson DG, Arnold JE, Tanenhaus MK. Tic Tac Toe: effects of predictability and importance on acoustic prominence in language production. *Cognition*. 2008; 106(3):1548–1557.10.1016/j.cognition.2007.06.009 [PubMed: 17697675]
- Watson, D.; Arnold, JE. Not just given and new. The effects of discourse and task based constraints on acoustic prominence; Poster presented at the 2005 CUNY Human Sentence Processing Conference; Tucson, AZ. 2005.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript