1 Awareness, Salience, and Stereotypes in Exemplar-Based Models of Speech Production and Perception

Katie Drager and M. Joelle Kirtley

The contributions to this volume discuss the degree to which awareness plays a role in how language is produced (Babel, Zimman), acquired (Nycz), and processed (Beck, Carmichael, Squires). The conclusions underscore the need for models of speech production and perception that can account for different amounts of awareness and attention. In this chapter, we seek to lay out how awareness, salience, and stereotypes are implemented within exemplar-based models of speech production and perception, reflecting on predictions that these models make in regard to awareness and salience.

Before addressing how awareness and salience might work within exemplar-based models of speech, let's talk briefly about awareness and how the term is used. Sometimes researchers use the term to refer to an awareness of a social category (e.g. Jock) or a linguistic variant (e.g. fishin'), and other times they refer to the awareness of a relationship between a social category and a linguistic variant (e.g. Midwesterners say 'pop'). These differences matter when considering if/how awareness is represented in the mind and when thinking about the cognitive processes through which awareness might influence speech. In this chapter, we focus on the last of these types of awareness: the awareness of a sociolinguistic variable.

But what do we even mean by awareness? Awareness is usually taken to mean one's consciousness of events or experiences. Some prior instance of noticing is required for awareness, but deliberate effort and instruction are not. In line with Squires (this volume) and others (Bowers 1984; Schmidt 1990), we differentiate between noticing a difference (which leads to awareness) and perceiving a difference (which, in the absence of noticing, does not). Perception without awareness is possible because many cognitive processes are automatic or reflexive, which contrast with processes that are controlled or reflective (Lieberman 2003; Schneider and Shiffrin 1977; Shastri and Ajjanagadde 1993). As Lieberman explains, "controlled processes . . . typically involve some combination of effort, intention, and awareness, tend to interfere with one another, and are usually experienced as self-generated thoughts. Automatic processes . . . typically lack effort, intention, or awareness, tend not to interfere with one another, and are usually experienced as perceptions or feelings"

(Lieberman 2003: 44). Controlled and automatic processes differ, and they have varying effects on an individual's linguistic behavior.[1] We should, therefore, expect different behavior depending on a speaker's or hearer's attention.

The necessity of the distinction between controlled processes and automatic processes in sociolinguistics finds support from the different behavior observed for markers and stereotypes in speech production (Labov 1972; Trudgill 1981)[2] and across different experimental tasks that vary in the degree of introspection they require (Hay *et al.* 2010). Both controlled and automatic processes contribute to sociolinguistic behavior and, therefore, should be accounted for within our models of speech production and perception.

In this chapter we discuss automatic processes and explore the treatment of awareness in exemplar-based models of speech production and perception. We first present what we mean by exemplar-based models. Next, we discuss how attention, salience, and stereotypes are accounted for in these models, and we step through the results from our previous work in order to make explicit how the models predict the observed behavior.

What Is Exemplar Theory?

Exemplar Theory is a collection of cognitive models in which experiences are encoded in the mind as episodic memories, known as exemplars. The models are often (but not always) implemented computationally to test the model's predictions (e.g. Hintzman and Ludlam 1980). Exemplar models originated in psychology (Brooks 1978; Hintzman and Ludlam 1980; Schacter *et al.* 1978) and continue to be influential (e.g. Nosofsky *et al.* 2011). Highly relevant to sociolinguistics are exemplar models from social psychology that examine the perception of people. In these models, individuals have mental representations of categories which contain numerous types of information about the category; including stereotypes and beliefs about the category, values associated with the category, one's past interactions with people associated with the category, and specific category exemplars (Bern 1972; Eagly *et al.* 1994; Haddock *et al.* 1993; Nosofsky *et al.* 1994; Smith 1998). The activation of exemplars affects individual behavior and attitudes, often with no awareness of the activation of the exemplars or attitude (Lewicki 1986; Smith and Zárate 1992).

Linguists have extended and modified exemplar-based models to explain linguistic behavior (Johnson 1997; Pierrehumbert 2001). In such models, when

[1] It is worth noting that behavior, such as a stroke in a tennis match, can result from a combination of controlled and automatic processes. "The conscious decision about which stroke to attempt may be the result of a controlled process, whereas the actual stroking of the ball may be automatic" (Fazio 1990: 97).

[2] Labov's treatment of indicators, markers, and stereotypes is discussed in several contributions to this volume (e.g. Carmichael), so – in the interest of space – is not discussed further here.

a listener encounters someone saying *pizza*, the memory of that utterance is stored as its own representation (i.e. an exemplar) that is distinct from representations that encode other occasions when the listener heard the word *pizza*, even when those utterances were produced by the same speaker.[3] But there is clearly something similar about all of the different representations of *pizza*; they are not floating around in the mind, unrelated to one another, but instead, form what is called an exemplar cloud. Exemplar clouds are often thought of at the word level (Johnson 2005; Wedel 2006), but in the multidimensional space that is the mind, they can occur simultaneously at segmental and lexical levels (Pierrehumbert 2001), and potentially at other levels of the grammar as well.

Some researchers who work with exemplar-based models implement models that are strictly episodic (what Sherman refers to as "pure exemplar models" (1996: 1127)). Proponents of these types of exemplar models argue that any generalization that takes place occurs online without levels of representation beyond the episodic memories (Hintzman 1986; Nosofsky 1986). However, exemplar-based models of speech production and perception most commonly incorporate both episodic memories and generalized levels of representation (Goldinger 2007; Hay *et al.* 2013; Hay *et al.* 2006b; Johnson 2007; McLennan 2007; Nielsen 2011; Pierrehumbert 2006; Sherman 1996). The generalized levels are deemed to be necessary because individuals can produce and perceive words that they have not encountered before, and listeners generalize learned knowledge within natural classes and phonemes (McQueen *et al.* 2006; Norris *et al.* 2003). At the same time, episodic memories are seen to be necessary because listeners are sensitive to speaker- and utterance-specific phonetic detail (Craik and Kirsner 1974; Goldinger 1997; Palmeri *et al.* 1993), as well as phonetic cues related to affect (Gobl and Ní Chasaide 2003; Morton and Trehub 2001; Nygaard and Lunders 2002). For example, in a task where listeners identify whether or not they previously heard a word, they are more accurate when the same voice is used between trials than when it is a different voice (Craik and Kirsner 1974; Palmeri *et al.* 1993), and they remain sensitive to the speaker-specific phonetic realizations for at least a week (Goldinger 1997). Additionally, some researchers argue that sound change is linked with token frequency (Bybee 2001, 2002); such findings are consistent with exemplar-based models because frequency effects are a product of having episodic representations. It is important to note that these results are also consistent with other experience-based models that store information about the relationship between phonetic realizations and token frequency. However, because the relationship can be computed online

[3] People who talk about cognitive models often talk about representations of the word *cat*. We decided to mix things up and talk about *pizza* instead. Maybe we're hungry.

in an exemplar-based model (thus avoiding the need for explicit storage of the association), an exemplar-based account is an especially elegant solution.

Because of results such as those outlined above (e.g. Palmeri *et al.* 1993), it is posited that a great deal of phonetic detail is encoded in exemplars. This phonetic detail may include any number of phonetic traits, including voice onset time of plosives, nasalization of vowels, or the center of gravity of fricatives. Thus, as shown in Figure 1.1, the word *pizza* that our listener heard earlier would be stored in precise acoustic detail: encoding, for example, the duration of the aspiration in /p/, the vowel quality of /i/, and the center of gravity of the aperiodic energy in /s/. For segmental phonetic cues, each phoneme that was ultimately identified is indexed to the phonetically rich exemplar. In this chapter, we refer to the phonetically detailed representations as phonetic exemplars.

Phonetic exemplars are the main focus of our discussion since most sociolinguistic work that discusses exemplar models has focused on phone variation. However, exemplar models within variationist linguistics are not limited to phonetics, phonology, or the lexicon; there are, for example, episodic models of syntactic variation (Abbot-Smith and Tomasello 2006; Bybee and Cacoullos 2008; Erker and Guy 2012). There is also evidence that phonetic

Figure 1.1 Exemplar Cloud of the Word *Pizza*

detail is encoded in the lemma (Drager 2010, 2011a; Gahl 2008; Plug 2005), across word boundaries (Hay and Maclagan 2010), in chunks of speech that constitute two or more words (Bybee 2002), and in association with certain morpho-syntactic representations (Walker 2008). Much more work is needed to clarify the relationship of probabilistic information between different levels of the grammar.

In addition to the storage of language-specific information, other kinds of information are stored. This information includes any number of markers of an individual's identity, including broad social categories (e.g. lesbian), membership in a community of practice (e.g. Norteña), or a social practice itself (e.g. wears red heels). Because these markers include a wide range of different kinds of information and the different kinds of information are, in turn, indexed to each other, we use the intentionally vague terms "social meaning" and "social indices," and we wish to make explicit that a great deal of information is stored, detailed information that goes well beyond what might be considered traditionally linguistic information. While the term "information" may be more apt since divorcing the linguistic from the social is artificial, we feel it necessary to emphasize the presence of social information in the models given the history of linguistics and the tendency of some researchers to downplay or disregard the role of the social. In these models, socially meaningful information is indexed to linguistic exemplars, and this indexing may be direct or indirect depending on the individuals' experiences. Social indexing occurs automatically, without conscious effort by the perceiver.[4] "Thus, the exemplar model intrinsically captures the observation ... that no natural human utterance offers linguistic information without simultaneously indexing some social factor" (Foulkes and Docherty 2006: 426). An exemplar-based model with social indices not only accounts for sociolinguistic variation, it is a linguistic theory which predicts that socially conditioned variation will exist. Because utterances are stored as individual memories and these memories encode and are indexed to a great deal of information and because speech perception and production rely on these stored exemplars, variation is a natural consequence of the model.

How Does Speech Perception Work?

Listeners encounter utterances, and these utterances are then stored. Once stored, these representations can be activated. Incoming speech activates the

[4] McGowan (this volume) contends that awareness is required for social indexing in exemplar-based models, but we propose that his results can be explained through perceived social information (such as perceiving the speaker as Asian or Chinese when in fact the speaker is Japanese) and through the storage of imitated speech, as outlined in this paper.

exemplars it is most similar to, and perception is biased towards the activated exemplars.[5] Thus, an incoming [p] activates exemplars of [p] more than, say, [s], [k], or even [pʰ], biasing perception towards [p]. The more closely the realization of the incoming [p] matches an exemplar, the more likely the exemplar will be activated. Other exemplars are also activated, particularly where there is a great deal of allophonic variation (Boomershine *et al.* 2005; Johnson 2006).

Activation of exemplars is gradient: one exemplar of *pizza* can be more or less activated than another exemplar of *pizza*, and those representations that are activated the most influence perception the most (Hintzman 1984). Activation spreads to exemplars that are indexed to already activated exemplars, and exemplars that are recently and frequently activated reach full activation the fastest and therefore bias perception the most. While exemplars decay over time, activation slows decay (Lacerda 1995), so exemplars that encode frequently encountered realizations resist decay.

Because stored social information is indexed to detailed linguistic information, activating linguistic exemplars activates the social information to which those exemplars are indexed. Thus, if our listener is at a café and overhears someone behind her say something about "eatin' pizza," she may infer certain things about the speaker based on people she has encountered before, as well as things she's heard about groups of people and the way they supposedly talk. A large amount of literature exists, both within linguistics and social psychology, demonstrating that listeners make judgments about speakers based on their speech, and these judgments are highly consistent across different listeners (Addington 1968; Aronovitch 1976; Harms 1961; Kirtley 2010; Carmichael, this volume). In an exemplar model, this occurs because linguistic exemplars are activated upon perception, which in turn activates associated social exemplars. This process happens automatically and, therefore, awareness of the specific linguistic variants or their association with social categories or traits is not necessary.

However, the process is not as simple as an incoming utterance activating a phonetic exemplar which in turn activates one and only one social meaning; we know that the social meaning of a linguistic variant shifts depending on contextual factors, including characteristics attributed to the speaker (Campbell-Kibler 2007) and other linguistic cues in the signal (Levon 2011). In the exemplar-based model proposed by Drager (2009), patterns of activated exemplars are indexed to personal styles (Drager 2009: 184) and may, in some cases, make up the styles themselves, in which case there is no abstract representation of the style. These patterns of activation may be over any

[5] In Nosofsky's (1992) model, activation relies on a combination of the exemplar's strength in memory, its similarity to the incoming utterance, and random noise (Nosofsky 1992: 386).

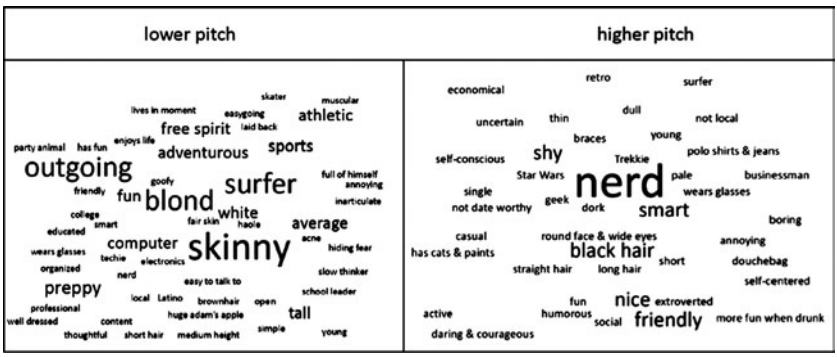

Figure 1.2 Perceived Traits for a Male Voice with Lower Pitch and Higher Pitch

number of stylistic components, including different linguistic variants. If the speaker who was talking about "eatin' pizza" also produces monophthongal /ɑi/, we expect our listener to attribute different social characteristics upon hearing "eatin' pizza" than were the speaker to produce diphthongal /ɑi/ (cf. Campbell-Kibler 2007). According to the exemplar-based model proposed by Drager (2009), this happens as a result of activation patterns that include these realizations, and social characteristics such as "educated" are primed via the activated style. The model also predicts that activation can spread to and/or from other stylistic components (e.g. hairstyle) that the listener associates with the style.

To help demonstrate how components of styles are activated in this model, Figure 1.2 shows two tag clouds from a matched guise experiment during which American-English-speaking participants responded to the open-ended question: *What do you think this speaker is like?*[6] Across the two guises, voice and content were controlled and mean pitch was manipulated. Responses are shown in Figure 1.2, with larger text indicating a larger number of participants who responded with that word or phrase. The difference in pitch seems to be related to a difference in style, ranging from what might be called a "sporty professional" style in the lower pitch guise to a "nice nerd" style in the higher pitch guise. Many participants who took part in this experiment responded with traits associated with a style rather than a label for the style itself. In the model outlined above, this happens when the incoming utterance activates phonetic exemplars to which it is most similar. Because pitch is included in the phonetic exemplars, exemplars that closely match in pitch are activated, and those that encode other phonetic

[6] This experiment was reported by Drager *et al.* (2010b).

information similar to that found in the incoming utterance are activated the most. The activation then spreads from the activated phonetic exemplars to indexed social information. The styles are perceived within the context of other social information attributed to the speaker (e.g. male, heterosexual) because this information, too, is activated and can serve to spread activation to social information with which it is indexed.

Exemplar-based models with social indexing also predict that listeners' perception of linguistic variants will be biased as a result of contextual factors. For example, our listener's surroundings – including the visible clientele of the café – influence her expectations about the speech she's about to hear. If our listener sees a stranger, the stored social exemplars of people who look most similar to that stranger are activated and her perception would be biased as a result. In fact, there is mounting evidence to support the prediction that social information attributed to a speaker influences how a listener will perceive their speech (Hay *et al.* 2006a; Hay *et al.* 2006b; Koops *et al.* 2008; Niedzielski 1999; Strand 1999). So, if the stranger in the café produces a linguistic variant that differs from the linguistic exemplars that are indexed to the activated social exemplars, our listener's speech processing will be slower than if it is similar (cf. Staum Casasanto 2008), and if the stranger produces an ambiguous word (e.g. does the cat sleep in the *litterbox* or *letterbox*?[7]) the listener's perception will be biased towards the speech of people deemed to be similar to the stranger (cf. Hay *et al.* 2006a). The evidence suggests that the flow of activation goes both ways: from the linguistic to the social and from the social to the linguistic. This is important to consider when thinking about how linguistic information is stored in the mind. Most models of speech perception would not predict these results because processing is strictly bottom-up (e.g. Klatt 1979) or because social information is not considered in the models (e.g. McClelland and Elman 1986).

How Does Speech Production Work?

Stored exemplars also influence speech production. Activated linguistic exemplars bias production towards the variants they encode. As with perception, activation is fastest for recently and frequently activated exemplars, and social and contextual information can activate phonetic exemplars to which it is indexed. This can account for a wide variety of work in variationist sociolinguistics, including effects of topic (Rickford and McNair-Knox 1994), context

[7] Katie once misinterpreted a friend from New Zealand as saying that her cat sleeps in the litterbox. The priming of social information failed Katie that day, probably because other primes (i.e. cat → litterbox) were stronger than the social primes, and *letterbox* is not a lexical item that is found in Katie's native dialect.

(Bell 1984; Coupland 1980; Podesva 2008), and stance-taking (Johnstone 2009; Kiesling 2009).

Of course, attitudes also play a role in both speech production and perception. For example, a speaker's attitude towards an interlocutor influences the direction and amount of speech accommodation (Babel 2010; Bourhis and Giles 1977; Giles 1973; Giles *et al.* 1991). In an exemplar-based model, positive attitudes towards a person or group may activate linguistic exemplars associated with that person or group, resulting in speech convergence (Drager *et al.* 2010a: 31).[8] To account for effects of negative attitudes, Drager *et al.* (2010) offer two possibilities. The first is that negative attitudes towards a social group could influence production and perception by activating alternative social exemplars that encode social information about which the speaker-hearer has positive associations. The second explanation is that activation is inhibited by negative attitudes. In models that allow inhibition (e.g. Pierrehumbert 2001), production is biased away from variants encoded by the inhibited exemplars. Teasing apart these two interpretations remains a task for future work, and it is quite possible that both mechanisms exist and can occur simultaneously or are relied upon at different times, such as during different kinds of tasks.

Taken together, the exemplar-based models presented thus far predict that individuals will have so-called "knowledge" of sociolinguistic variation that they are not aware of. If speech is stored in the mind as socially indexed and phonetically rich episodic memories, an individual's speech production and perception can be influenced by patterns present across the exemplars even when they don't notice the patterns.[9] This contrasts with Labov's argument that variables below a speaker's conscious awareness "could hardly ... be the direct objects of social affect" (Labov 1972: 40). This conclusion is unavoidable if we assume that to be socially meaningful or have social effect is, by definition, an awareness of a relationship between social factors and language use. However, we argue that individuals do not need to be aware of variation in order for that variation to be socially meaningful. In the model outlined above, stored social information can be activated during the construction of a speaker's identity: a speaker activates social representations (e.g. intelligent, laidback, feminine) and, by doing so, activation inadvertently spreads to indexed linguistic information. It is possible, therefore, that production can be biased towards certain linguistic variants and that the bias can be socially motivated, while the speaker remains unaware of the association.

[8] How (or even whether) attitudes are stored remains underspecified in exemplar-based models.
[9] Docherty and Foulkes refer to associations that arise in this way as awareness, albeit implicit (Docherty and Foulkes 2014: 47). This contrasts with the treatment of awareness in this chapter which necessitates consciousness.

In order to demonstrate how sociolinguistic variation arises in an exemplar-based model even when speakers are unaware of the sociolinguistic variation, we briefly step through results from a study conducted at an all girls' high school, referred to as Selwyn Girls' High. The production data consist of recorded conversations and are reported in more detail in Drager (2011a). While there were many different social cliques observed at the school, the cliques were categorized depending on whether or not they ate lunch in the Common Room. Girls in Common Room groups (e.g. The PCs) set and perpetuated the school's norms of dress, behavior, and beliefs, and girls in most of the non-Common Room groups (e.g. The Goths) actively rejected these norms. The results indicate a number of phonetic differences across the social groups and the functions of *like*; here we focus on realizations of quotative *like* (e.g. I was like "turn that stupid thing off!") across Common Room and non-Common Room groups.

Common Room girls produced longer /l/ durations and more monophthongal vowels in quotative *like* than did non-Common Room girls. Although the girls were not aware of the phonetic differences, the differences are believed to be a result of identity construction (Drager 2011a). This is possible in an exemplar-based model with social indexing because identity construction occurs through activating social exemplars, and activation spreads from social exemplars to phonetic exemplars. Activation of social exemplars is related to an individual's attitudes towards social groups and characteristics. As examples, we discuss two speakers from two different non-Common Room groups: Holly, a member of Sonia's Group who idealized members of The PCs (a popular and powerful Common Room group), and Santra, a member of The Goths who consciously and outspokenly took part in practices that differed from girls in Common Room groups.[10] Holly produced realizations of quotative *like* that were similar to those produced by Common Room girls, whereas Santra produced realizations that were very different from those produced by Common Room girls (Drager and Hay 2012). In an exemplar-based model, the difference in realizations arose because of the social representations that were activated when Holly and Santra spoke; social representations associated with The PCs were activated when Holly talked because of Holly's alignment with The PCs, and activation then spread to indexed phonetic exemplars.[11]

[10] Sonia's Group was classified as a non-Common Room group because they did not eat lunch in the Common Room. However, their style of dress, weekend activities, and topics of conversation were similar to those of Common Room groups.

[11] For simplicity, we discuss the activation of exemplars at the social group level (e.g. The PCs) rather than social characteristics (e.g. bossy) or individual speakers, and we treat attitudes as categorical (i.e. positive or negative). Of course, the process of identity construction (and therefore the activation of social exemplars) is more complicated than this implies. Simultaneous activation of a wide variety of possibly conflicting information is possible in an exemplar-based model.