Abstraction-based efficiency in the lexicon

ANNE CUTLER

Max Planck Institute for Psycholinguistics
Radboud University Nijmegen
University of Western Sydney

Abstract

Listeners learn from their past experience of listening to spoken words, and use this learning to maximise the efficiency of future word recognition. This paper summarises evidence that the facilitatory effects of drawing on past experience are mediated by abstraction, enabling learning to be generalised across new words and new listening situations. Phoneme category retuning, which allows adaptation to speaker-specific articulatory characteristics, is generalised on the basis of relatively brief experience to words previously unheard from that speaker. Abstract knowledge of prosodic regularities is applied to recognition even of novel words for which these regularities were violated. Prosodic word-boundary regularities drive segmentation of speech into words independently of the membership of the lexical candidate set resulting from the segmentation operation. Each of these different cases illustrates how abstraction from past listening experience has contributed to the efficiency of lexical recognition.

The lexicon is where sounds meet meaning, and from the listener's point of view the lexicon is the core of the speech comprehension system. Working out what a speaker has said largely consists of identifying the words that have been spoken. Our subjective experience is that this is not a hard task at all: lexical processing is, above all else, outstandingly efficient. This is not just a function of robust and efficient front-end processing – though listening to the native language certainly involves that too. The argument of this paper is that lexical efficiency is at least in good part founded on listeners' ability to abstract from and generalise their word recognition experience so far, to facilitate word recognition in the future. Abstraction and generalisation form the heart of human cognitive efficiency, in speech processing as in other mental activities.

To the members of the Laboratory Phonology community, it is now a truism that language users are sensitive to the statistical distributions comprising their linguistic experience, and tailor speech perception and production decisions accordingly. In early Laboratory Phonology conferences a recurring theme, for instance, was

JLP 1 (2010), 301–318
DOI 10.1515/LABPHON.2010.016

1836-6346/10/01–301

the interaction of usage patterns with language change, producing asymmetric application of sound changes across the lexicon (see, for instance, Yaeger-Dror, 1994; Bybee, 2000; Beckman and Pierrehumbert, 2003). The rise of exemplar theory in linguistics rapidly found a reflection in the contributions to the conference series (see, e.g., Pierrehumbert, 2002; Jurafsky, Bell and Girand, 2002), and indeed, by the time of the tenth meeting, became a central theme (Fougeron, Kühnert, d'Imperio and Vallée, 2010). Exemplar theory, developed in psychology as a model of memory (e.g., Nosofsky, 1986; Hintzman, 1986), has become in linguistics a popular tool for capturing the role of subtle statistical variation in linguistic performance.

The most powerful way to put statistics to use, however, is to base generalisations upon them. This is one of the ways in which abstraction-based efficiency will be illustrated in the present paper. The first example, however, concerns the flexibility of the categories in terms of which lexical processing is organised.

1. Category flexibility

It would be very unwise of me to ignore a stoplight in Australia or New Zealand just because its shade of red was not exactly that of the stoplights where I live, in Europe. Red is red; a member of the category red instantiated as a traffic signal communicates a clear message. The message does not depend upon the colour's precise realisation in the particular stoplight, but upon its categorisation.

Categorisation supports the efficient operation of all aspects of cognitive processing, from lexical access (at the end of the continuum closest to our present concerns) to visual processing (at what may seem to be a rather distant other end). Colours are fundamental categories of visual processing. Nevertheless, colour interpretation is highly context-dependent; in the absence of any external reference other than the immediate context, exactly the same shade may be categorised as "yellow" if it is in shadow and surrounded by dark colours, but as "brown" if it is brightly lit and surrounded by light colours. External reference, however, can alert the viewer to the necessity to adjust categorisation for the context (this is one of several reasons why traffic lights are redundantly coded: red is always on top). We know that a piece of paper is white whether we view it in a dimly lit room or in bright sunlight, even though the two cases present our eyes with different sensory information.

A recent study by Mitterer and De Ruiter (2008) proved that the boundaries of colour categories can be adjusted if reference to stored knowledge indicates that they should be. In a two-part experiment, participants first performed an object recall task. Among the objects they were required to remember might be, for example, a banana (appropriately, bright yellow), and a carrot (unusually pale in colour). Another group of participants might then be presented with an unusually darkish banana and a classically orange-coloured carrot. The shade which, in this presentation, was unusually pale for a carrot and unusually dark for a banana was,

crucially, the same colour; the first group was effectively trained to categorise this colour as orange (because it was the colour of a carrot), while the second group was trained to categorise it as yellow (because it was the colour of a banana). In the second part of the study, the participants then performed an apparently unrelated task; they judged the colour, which varied along a yellow-to-orange continuum, of a series of single socks.

The two groups' colour categorisation judgements differed. The group that had been trained to include the very unusual orange in their orange category produced overall more "orange" judgements, while the group that had been trained to include that same shade in their yellow category produced overall more "yellow" responses. The adjustment of their category boundaries was apparent along the continuum, not just at the most ambiguous point (i.e., the exact shade used in the object-recall part of the study). Clearly, their stored knowledge of the colour category appropriate for fruits and vegetables had influenced their assessment of what (in this experimental situation, at least) a given colour should look like. Category decisions were modified to suit the situation indicated by the input received.

Mitterer and De Ruiter's experiment was directly modelled on an investigation of phoneme categorisation by Norris, McQueen and Cutler (2003). In that study, the first part of the experiment was an auditory lexical decision task, in which listeners just had to decide whether spoken items were real words or not. For one group of participants, 20 words in the experiment contained a sound which had been established in a pre-test to be ambiguous between /f/ and /s/; this sound replaced what should have been /s/. For instance, the sound might have replaced the final phoneme of *horse* – though in fact the original experiment was in Dutch. This group also heard 20 words (such as *giraffe*) with a normal /f/. For another group, 20 words such as *giraffe* which should have contained /f/ had the ambiguous [f-s] instead, while the 20 /s/-words such as *horse* contained a normal /s/. There were no other instances of /s/ or /f/ anywhere in the 200-item list, in words or in nonwords. In this task, the tokens containing the modified sound were generally accepted (i.e., they were taken to be words; although responses were slower than to words with normal pronunciation). The acceptance indicates that the ambiguous sound was interpreted as an instance of the sound proper to the word in question. Crucially, in the second part of Norris et al.'s study, a phoneme categorisation task revealed that those listeners who had heard the modified [f-s] replacing /f/ had expanded their /f/ category, while the listeners in the other group, who had heard the same sound replacing /s/, had expanded their /s/ category. As in the colour study, the boundary adjustment was apparent along the continuum, not only at the most ambiguous point which was identical to the [f-s] sound they had actually heard.

The same flexibility can be observed with printed-letter categories. An ambiguous letter form midway between H and N, presented in lexical contexts such as WEIG- or -OIST in a visual lexical decision experiment, should be interpreted as H; in lexical contexts such as REIG- or -OISE, in contrast, participants should interpret it as N. Norris, Butterfield, McQueen and Cutler (2006) tested this, in an

experiment again modelled on the 2003 study with phonemes. In their experimental design, the initial (visual) lexical decision experiment was followed by a letter categorisation experiment, in which the participants judged whether a letter form was more like H or N. Here the first group's H category proved to have expanded, while the second group's N category had expanded, and the parallels continued to hold in that once again the expansion was observed along the ambiguity continuum.

Control conditions in all these similarly designed studies established that participants needed a reference against which to establish an identity for the unusual category exemplar. Exposure alone was not sufficient, and nor was contrast. In the phoneme study, it was not sufficient to hear one of the clear fricatives alone and no ambiguous sounds, or the ambiguous sound in nonwords only. In the letter study, it was not sufficient to encounter the ambiguous letter in contexts such as MEIG- or -IST which would produce a nonword irrespective of the chosen letter interpretation (MEIGN or MEIGH or HIST or NIST – they are all pronounceable in English, but none of them are existing words). In the colour study, it was likewise insufficient to see one of the canonical colours alone and the intermediate colour only in a context unspecified for colour, such as on a car or a dress. If the boundaries between categories are to be readjusted, it is imperative that the processor be presented with clear evidence of the particular category to which an anomalous exemplar should be assigned. This evidence can be lexical identity (as it was in the phoneme and letter studies), or it can be an object label (as in the colour study); what is important is that the evidence force the interpretation of the ambiguous token.

This kind of category retuning process is extremely robust, and perceivers' sensitivity to relevant evidence is quite subtle. The retuning of phonemes can be induced, for instance, by evidence presented in pseudowords, if one of the endpoint interpretations would violate constraints on the viability of phoneme sequences. An ambiguous [f-s] token is thus interpreted as [f] before [r] but as [s] before [n] (*frulic* and *snuter* rather than **srulic* and **fnuter*), and a subsequent phoneme categorisation task again reveals a boundary shift (Cutler, McQueen, Butterfield and Norris, 2008). Visual cues to articulation can also provide evidence which motivates category retuning (Bertelson, Vroomen and De Gelder, 2003; Van Linden and Vroomen, 2007). Retuning appears in vowels (Maye, Aslin and Tanenhaus, 2008), and in consonants of different types (Kraljic and Samuel, 2006, 2007; Sjerps and McQueen, 2010), as well as in other categorisations that are similarly relevant for the establishment of lexical identity – thus in Mandarin, the boundaries between lexical tone categories can be made to shift in the same way (Mitterer, Chen and Zhou, in press).

2. Categorical decision-making

From the evidence reviewed in the preceding section we can conclude that the flexibility of category boundaries is a general cognitive effect, designed to serve the

purposes of the important mental operation that we call categorisation. It is not speech-specific, or language-specific. In all the domains in which it manifests itself, speech included, it ensures that the right category is selected irrespective of fluctuations in the sensory input caused by varying environmental conditions. Here the speaker who utters a particular phoneme, for example, can be viewed as an environmental dependency, not effectively different from variables such as the font in which a text is printed, or the angle at which light falls on a coloured object such as a traffic light. Such environmental variability is not in principle relevant to the message (in speech, in printed text, or in communicative use of colour); it needs to be factored out. In the speech perception literature, this process is usually known as normalisation. Normalisation is thus by no means a speech-specific process (indeed, Broadbent, Ladefoged and Lawrence, 1956, pointed out the parallel with colour perception in their important study of vowel normalisation, using the white-paper example borrowed above). Categorisation underlies cognitive efficiency in all aspects of perception, and adjustment of category decisions to account for contextual factors is an operation found in all cognitive domains, not just in speech.

It is important, in providing an account of the role of categorisation in cognition, to distinguish between the implicit categorical decisions that inform the recognition of speech (or indeed of other visual or auditory experience), and the explicit categorical decisions that may be required in experiments, including many of those reviewed above. An implicit categorical decision is involved when a speech input *diffi-* continues as *diffic-* rather than as *diffid-* or anything else; the effect is that the word *difficult* will receive further support from the signal, and its recognition will become more probable, while alternative lexical candidates such as *diffident* will become less probable. The point at which *difficult* diverges from *diffident* is the fifth phoneme in the input. But this does not entail that the speech processor has made an explicit choice between /k/ and /d/, or has recoded the input in any form that involves a separate representation of the phoneme /k/. Indeed, the gradient activation of competing words resulting from listeners' exploitation of coarticulatory information provides evidence against such explicit choices (Marslen-Wilson and Warren, 1994; Dahan, Magnuson, Tanenhaus and Hogan, 2001). Explicit choice is however indeed required in a categorisation task (involving phonetic or any other type of category decisions), and also in related tasks such as same-different judgement or target detection.

In making explicit categorical decisions, people efficiently and accurately combine relevant information from multiple sources. The Merge model of decision-making in speech (Norris, McQueen and Cutler, 2000) provides an account of this process, and the account is based on the separation of explicit from implicit categorisation. All sorts of knowledge can be pertinent to a given decision. Of obvious relevance are talker characteristics such as vocal tract size, and environmental factors presenting masking or affecting signal transmission (see, e.g. Ohala, 1995, for a signal transmission effect on vowel categorisation similar to that reported for

talker characteristics by Broadbent et al., 1956). But lexical factors can also influence listeners' decisions about what phonemes they are hearing (e.g., Ganong, 1980) or their speed of detecting a phoneme target (e.g., Connine, Titone, Deelman and Blasko, 1997), and so can the plausibility of a larger sentence context (e.g., Borsky, Tuller and Shapiro, 1998; Morton and Long, 1976). Some effects are to a greater or lesser degree indirect. Thus the rate of speech implied by just a single long or short vowel can influence decisions about the identity of an immediately preceding consonant (Miller and Liberman, 1979), a photograph of a person can suffice to cue the relevant talker characteristics (Hay, Warren and Drager, 2006), and exposure to a dialect label can bias the recall of the nature of a vowel sound (Hay, Nolan and Drager, 2006). Combined use of very different information sources – acoustics of the signal, stored knowledge in the lexicon or general knowledge in memory, abstract labels and prior episodes of cognitive or sensory processing – in an integrated fashion is however no problem for a system such as that proposed by the Merge model. The model output is explicit decisions, and in line with a fundamental postulate of signal detection theory, each of these decisions can be biased by any number of disparate factors.

All of these factors will also be useful in informing the implicit decision-making in speech perception, but if there is no separate representation of phonemes in that process, then there is no requirement for all disparate factors to operate in the same manner and at the same level of processing. It follows that categorisation tasks do not necessarily provide evidence for exactly how any given factor impinges on the processing of speech (only that it can impinge!). Converging evidence from tasks that tap into speech recognition is needed, as the following sections illustrate.

3. Phoneme retuning and its efficiency

The retuning of phoneme category boundaries as first observed by Norris et al. (2003) is in the first instance efficient because it allows listeners to adjust to input from different talkers (see Cutler, Eisner, McQueen and Norris, 2010, for further elaboration of this argument). As would be predicted from such a claim, the retuning proves to be both talker-specific and segment-specific (Eisner and McQueen 2005; though see Kraljic and Samuel, 2006, 2007, for discussion of when and how generalisation across broader segmental classes should occur). The retuning is also stable across time (Eisner and McQueen, 2006; Kraljic and Samuel, 2005). But it does not always happen; it is inhibited if reference to external information counter-indicates its reliability.

For instance, listeners do not retune phoneme categories when they hear an unusual (though contextually interpretable) sound if it is uttered by a talker who previously produced impeccably standard utterances of the same phoneme. Nor do listeners retune if they hear an unusual version of a sound which happens to co-occur with visual information that the speaker has a pen in her mouth at the mo-

ment the sound was uttered (Kraljic, Brennan and Samuel, 2008, for both these findings). Again, listeners do not retune if an unusual sound occurs in one specified phonetic environment, while the same sound uttered by the same talker in other environments has nothing unusual about it (Kraljic, Samuel and Brennan, 2008). The latter case could be reasonably interpreted as a dialectal feature; thus in each case, the retuning fails to occur when there is evidence that the form of the phoneme's realisation may be attributable to something other than the immutable configuration of a particular talker's vocal tract.

If there is insufficient information to guide interpretation of the ambiguous sound, retuning also fails to occur; for instance, an ambiguous sound in word-onset position is compatible with too many alternative lexical candidates for a rapid decision to be possible (Jesse and McQueen, submitted). Whereas *horse* is a word and *giraffe* is a word but *horf* and *girass* are not, rendering decision-making for phonemes at word offset quite easy, the same s/f ambiguity before a vowel such as /a/ could activate *farm, pharmacy, farther, sarcastic, sergeant, czar* and many more candidates. By the time disambiguating information arrives to motivate the choice for /s/ or /f/, the auditory trace of the ambiguous sound may no longer be useful. (Compare this case of a prevocalic onset consonant with the phonotactically possible vs. impossible onset clusters manipulated by Cutler et al., 2008; if a s/f ambiguity is immediately cleared up by a following /n/ versus /r/, the nature of the ambiguous sound is apparently still available, since in that case retuning does occur.)

The absence of retuning when there is insufficient information to motivate it (the singleton-onset case) is thus, one might argue, the other side of the coin from the absence of retuning when other information suggests that adjusting the boundaries between whole categories would not yield a reliable processing advantage (the pen-in-the-mouth case, etc.). The evidence always points to a benefit for listeners; wherever information about category boundary placement can be rapidly enough derived, and is apparently to a sufficient degree reliable, it will be incorporated into processing because doing so will facilitate future category identification.

An ingenious practical application of the retuning effect was demonstrated by Mitterer and McQueen (2009), who used it to train learners' facility in understanding dialectal varieties of a second language. It is a common experience of learners that even with high proficiency in a second language, speech in an unfamiliar variety can be hard to comprehend. The participants in Mitterer and McQueen's study were native speakers of Dutch with very good English, and they were exposed to movies in English. For one group, the English in the movie was spoken in broad Australian; for the other group, it was a marked Edinburgh Scottish variety. Some participants watched with subtitles in their native language (so they fully comprehended the meaning of what was going on), others watched with subtitles in English (so they realised what words were being spoken, even if some of those words were unknown to them), and others had no subtitles at all (so they were forced to rely on the speech input alone). After watching the movie, each group heard new audio input in each of the dialects, which they had to repeat.

All groups of course were better at repeating utterances in the dialect of the movie they had heard. However, the smallest same-dialect advantage was displayed by the group that had received Dutch subtitles; their improvement was less than that of the no-subtitles group. This suggests that with the meaning accessible in the visual text, the difficult speech had not been closely attended to. In contrast, the group that received English subtitles showed the greatest improvement of all. Mitterer and McQueen argued that knowing what words were being uttered enabled these listeners to deduce what phonemes were being uttered, and retune their phoneme category boundaries for the previously unfamiliar variety. The listeners attended to the visual text, but used it effectively to improve processing of the audio input too, which played out in a greater ability to make sense of new input encountered later. (The implications of this for listening training in a foreign language are rather obvious.)

4. Phoneme retuning and how it is generalised

Mitterer and McQueen's study showed that the phoneme retuning effect was of real use to listeners in comprehension. The learning that participants had derived from the movie had been generalised to the novel audio samples presented to them later. It is this generalisation that is the key to why phoneme boundary retuning is useful to listeners, and so rapidly and efficiently deployed by them in everyday life. The effects shown in the phoneme categorisation test phases of the Norris et al. (2003) study and its successors are thus not limited to explicit decision situations. Retuning would be of course of limited use if this were so; and it would likewise be of little use if it only applied to the particular tokens heard in training (in a single movie experience, or in the first phase of a two-part experiment, but also in the first words uttered by a new acquaintance). Its usefulness is predicated on its generalisability to future listening experience with the same talker, accent or dialect.

The generalisation has been demonstrated in several ways in the laboratory. Learning based on phoneme tokens occurring in a constant position in the word (for instance, in final position, where maximal information about phoneme identity should be expected to accrue: *horse, giraffe*) generalises to allow interpretation of the same phonemic ambiguity occurring in other positions in the word, such as word-medial (consider *muffle* versus *muscle*) or word-initial (*fleet, sleet*; note, though, that the relevant experiments were again actually carried out in Dutch: Jesse and McQueen, submitted). If appropriate, the learning can also generalise across speakers (Kraljic and Samuel, 2006, 2007, who found that while learning about fricative boundaries was speaker-specific, learning about stop boundaries was less so), and also across phonemes within a class (Kraljic and Samuel, 2006, again, who found that the generalisation of stop consonant voicing boundary retuning spread from an alveolar to a bilabial contrast). Kraljic and Samuel ac-

counted for the different results across manner classes with the argument that fricatives encode relatively more talker-specific information, and hence are used to adjust speech perception in a way most calculated to aid adaptation to new talkers, while the pronunciation of stops, being less informative about talker identity, is more likely to be mined for phoneme-class generalisations, such as whether a speech variety contrasts short-lag against long-lag stops, or prevoiced against lag.

Most obviously, generalisation is required across the lexicon. Learning how a phoneme is uttered in one word must bear implications for its articulatory form in other words of the language. This generalisation was first demonstrated by McQueen, Cutler and Norris (2006), whose two-stage phoneme retuning study first exposed listeners to an ambiguous fricative (s/f) in lexical decision stimuli, and then found that minimal pairs based on the same s/f contrast (such as *knife* versus *nice*, though once more the actual pairs were Dutch items such as *doof* 'deaf' versus *doos* 'box') were unambiguously interpreted in accord with the first-phase category assignment. The interpretation of a word cannot be measured via phonetic categorisation, of course, and nor will lexical decision serve the purpose, given that both members of such a pair are real words; the second phase of McQueen et al.'s study thus used a priming task to detect word interpretation via the presence or absence of facilitatory effects on later recognition events for the same word. Participants heard a prime word such as [naɪs/f] and saw on the screen in front of them a word or nonword on which they were required to make a lexical decision. For those who had heard the ambiguous s/f replacing /s/ in the earlier phase of the study, reponses to visually presented s-words such as NICE were more strongly facilitated than responses to visually presented f-words such as KNIFE. For those who had earlier heard the ambiguous s/f replacing /f/, in contrast, there was stronger facilitation for the visually presented targets that were f-members of the minimal pairs. Thus, the phoneme category boundary had been retuned in such a way that perception of all relevant words could benefit from it.

Such generalisation, beyond the words encountered in a first exposure phase, has been repeatedly established, and not only for the fricative contrast originally tested by Norris et al. (2003). The retuning of Mandarin tone categories induced in the study of Mitterer, Chen and Zhou (in press), for instance, also proved to generalise across words in the lexicon. Further, when listeners learn that a particular speaker substitutes a non-native phoneme for a native category, they not only learn how that sound should be identified in the first-phase exposure, as measured by their lexical decision accuracy, but they also interpret it unambiguously as the trained native category in a second-phase priming study like that of McQueen et al., indicating again that they have generalised the retuning to other words (Sjerps and McQueen, 2010).

The latter study also provided another important contribution to the overall picture of how phoneme retuning works. Sjerps and McQueen compared the strength of the priming effect in the case of a retuned category versus an original category. That is, in a priming study they compared the facilitation for NICE and KNIFE

displayed by (a) participants who heard an ambiguous prime such as [naɪs/f], where the identity of the ambiguous sound was dependent upon retuning induced in an earlier training phase, versus (b) participants who had not undergone any prior training, and just heard an unambiguous, normally articulated, *knife* or *nice*. There was no significant difference in the priming pattern across these two groups; it was strong, highly significant and similar in size in each case. This can only mean that a retuned category functions in exactly the same way, and just as effectively, as the same category before retuning. Retuning is a normal part of speech processing, allowing adaptation to new talkers and new accents in a way that is maximally conducive to efficient exploitation of the newly acquired articulatory information.

5. Abstraction of prosodic form

Abstract generalisation of lexical probabilities to new tokens is a regular feature of lexical processing. Indeed, it manifests itself in multiple ways. One such concerns the prosodic form of words: specifically, the well-attested generalisation that syllables uttered in isolation have a longer duration than the same syllables uttered as a part of words of two or more syllables. Listeners know this regularity, and exploit their knowledge of it in spoken-word recognition. Davis, Marslen-Wilson and Gaskell (2002) found that listeners' recognition of the visually presented word *doctor* was more effectively facilitated by prior priming with a short utterance of its initial syllable, while recognition of visually presented *dock* was better primed with a longer utterance of the same syllable. Likewise, Salverda et al. (2007) found that listeners hearing a short utterance of the syllable *cap* in an eyetracking study were more likely to look at a picture of a captain, while a longer utterance of the same syllable induced more looks to a picture of a cap.

The question about this prosodic knowledge is whether it is abstract knowledge generalised from past experience, or whether it is word-specific feedback from the lexicon given a token of *cap* or *dock*. To answer this question, Shatzman and McQueen (2006) taught listeners new words and carefully controlled the exact acoustic nature of their listening experience. First they taught them to name some nonsense objects, some of which formed pairs on the analogy of *dock* and *doctor* or *cap* and *captain*: *nim* and *nimsel* for example. They adjusted the recordings they presented in such a way that the *nim* in each of these items was actually the same length. So in every token of *nim* or *nimsel* that these listeners had heard, the *nim* had the same duration. The listeners then took part in an eyetracking study, in which they were asked to click on pictures of the same objects. Now the words were presented either with the same duration as before, or with the *nim* syllable lengthened or shortened.

If word-specific experience is the crucial factor, then the version which exactly reproduces previous experience should be the one which most effectively contacts

the learned representation. If abstract knowledge plays a role, though, then the longer version might be more likely to encourage looks to the object with the monosyllabic name, and the shorter version more likely to encourage looks to the object with the bisyllabic name. Shatzman and McQueen found that not only was abstract knowledge called upon, it dominated the actual experience: the object called *nimsel* was looked at faster if its first syllable was shorter than if it was long, or than if its duration was that used in learning. Likewise the object called *nim* was looked at fastest if its duration was lengthened.

So versions of words which have never previously been heard, but which code known prosodic regularities, are better matches to the stored representations than versions that match 100% of prior experience. Clearly listeners can call on abstract knowledge of prosodic patterning, including the tendency for polysyllabic names to have shorter first syllables. Moreover, they apply this knowledge to the recognition of words for which their experience has indicated an invariant duration. It is not only their word-specific experience which determines how the words are recognised. They exploit their knowledge about prosody, abstracted from experience with many different words, as well.

6. Abstraction of prosodic patterns in the vocabulary

Prosodic distributions in the vocabulary are also drawn upon for useful application of statistical generalisations; here, the segmentation of speech into its component words is the paradigm case. In the case of English and Dutch, the generalisation involves lexical stress; both languages have stress that is free, not fixed. But there is nonetheless a strong tendency in both languages for the primary stress of words to fall in a consistent location, on initial syllables, and an even stronger likelihood that initial syllables will not be weak, i.e., will bear primary or secondary stress. Computations carried out for English by Cutler and Carter (1987) on a real-speech corpus revealed 90% of the lexical vocabulary to be strong-initial, while equivalent analyses for Dutch by Schreuder and Baayen (1994) yielded a markedly similar figure of 87.5%.

Thus the strength of this distributional pattern is such that in both languages, it is helpful for listeners to assume that any syllable which is strong is very likely to be word-initial. This means that the pattern can be usefully exploited in segmenting continuous speech input; and abundant experimental evidence from both languages shows that listeners do indeed exploit it. In English, there is a significant tendency in natural slips of the ear involving word boundary misplacement for strong syllables to be interpreted as word-initial (*pledge allegiance* thought to be *led the pigeons*), but weak syllables to be interpreted as non-initial (*a must to avoid* interpreted as *a muscular boy*; Cutler and Butterfield, 1992); the same pattern can be elicited by presenting difficult-to-hear speech in the laboratory, both in English and in Dutch (Cutler and Butterfield, 1992; Vroomen, Van Zon and De Gelder,

1996). In the word-spotting task, which is a laboratory method for examining speech segmentation, words are harder to detect if detection requires overlooking a strong-syllable segmentation opportunity; *mint* is hard to find in *mintayf* (where the second syllable is strong, and hence putatively word-initial, suggesting a segmentation *min-tayf*). It is easier to find in *mintef* with a weak second syllable, suggesting no segmentation (Cutler and Norris, 1988). Again, this pattern is also observed in Dutch (Vroomen et al., 1996).

Importantly, the segmentation operation is separate from the process of lexical activation itself. Speech input, as it is processed by the listener, automatically makes available the words with which it might be compatible, and these word forms compete with one another until the correct sequence accounting for the input emerges (see McQueen, 2007, for an overview of the workings of spoken-word recognition). But the bias towards segmentation at strong syllable onsets does not just arise because more of the activated words begin with strong syllables. This can be easily seen by manipulating the number of competitors activated by the second syllable, as was done for English by Norris, McQueen and Cutler (1995) and for Dutch by Vroomen and De Gelder (1995). *Mintayf*, for instance, should activate a lot of competitors at the second syllable (*table, tailor, take, tame, taste* and many more). *Mintowf*, on the other hand, activates very few words beginning in the same way as its second syllable (*town, towel, tousle* and *tout*, and their morphological variants). The difference in size of the activated competitor set leads to a corresponding difference in the size of the inhibition in word-spotting – the more words compete for the second (strong) syllable, the more detection of the embedded word is slowed. But crucially, there is no effect of competitor set size when the embedded word is followed by a weak syllable. There are many more English words beginning with a /k/-initial weak syllable (all the words beginning *con-* and *com-*, for starters) than with a /t/-initial weak syllable (*today, tomorrow, taboo, toboggan* and not all that much else); but manipulating this difference in word spotting produces no significant effect on detection at all.

In other words, there is an effect of competitor set size when the second syllable is strong, because the statistically motivated segmentation operation induced by listeners' experience with strong syllables has occurred. There is no effect of competitor set size when the second syllable is weak, because in that case no segmentation has occurred. The competitor effects do not precipitate the segmentation; they are contingent upon it. The segmentation operation is driven purely by the application to the speech input of probabilities abstracted from the distribution of lexical experience.

7. Conclusion

Abstraction from listening experience induces processing heuristics on which the efficiency of speech comprehension is founded. Listeners adjust phoneme cate-

gory boundaries in a way appropriate for an individual speaker with deviant articulation; this adjustment, though clearly derived from episodic experience with the individual's utterances, is rapidly abstracted so that it can be applied to newly encountered tokens of the affected phoneme in different phonetic contexts, in different positions, in different words. The adjustment thereby facilitates future listening to the individual's speech. Adjustment is inhibited, however, if separately acquired knowledge indicates that a perceived categorical deviance is transient and not necessarily to be expected in future utterances. Nothing about this operation is blindly automatic; in contrast, it is intelligently tailored to the maximal benefit of listening efficiency.

The level of processing at which given knowledge sources are deployed in listening to speech depends on the level at which they are most useful to the processor. Lexical distributions do not affect segmentation willy-nilly; they are drawn upon to provide abstract probabilities which guide segmentation, and they are drawn upon separately to yield the specific probability set for lexical recognition once segmentation has been achieved. Implicit phoneme categorisation in lexical activation, similarly, is not under direct control of the lexicon, but is certainly aided by knowledge abstracted from a lifetime's experience with the patterning of the native language.

One such type of knowledge is phonotactics; this knowledge, we saw, could be applied to nonsense words which had never previously been encountered. Phonotactic constraints and phoneme sequence probabilities are lexically based, in that they are computed on the basis of the contents of a given vocabulary; knowledge of the probabilities is reflected in the direct correlation of well-formedness judgements about nonwords with the perceived likelihood of the form (Hay, Pierrehumbert and Beckman, 2003). However, phonotactic effects on processing are not derived online from the lexicon (for instance, lexical effects and phoneme sequence effects dissociate in phonetic categorisation: Pitt and McQueen, 1998). Likewise, phonotactic constraints and phoneme sequence probabilities are phonemically based, because they are distinct from lexical processing and because they are often language-specific; but unlike phonemic categories they are accessible to second-language learners and can be applied in parallel to native language constraints (Weber and Cutler, 2006). Despite the fact that phonotactic knowledge can be deployed in a gradient way, e.g., to provide estimates of relative well-formedness, it is the abstract knowledge of the probabilities that is drawn on to facilitate processing.

Prosodic probabilities are similarly extracted from listening experience, and similarly drawn on in speech processing. Monosyllabic words such as *cap* or *dock* are longer than their phonemic equivalents that form just a part of longer words such as *captain* or *doctor*; listeners' knowledge of this regularity leads them to appreciate the same distinction when it is available for distinguishing novel names such as *nim* versus *nimsel*, even in defiance of recent and exceptionless evidence that these particular novel items manifest no durational distinction. The probability

that word-initial syllables in English will be strong (and strong syllables will be word-initial) generates a segmentation heuristic which is applied to facilitate word recognition in continuous speech, and which indeed controls the set of candidates for recognition. In both these cases, listeners rely on knowledge abstracted from the totality of their past experience with the prosodic structure of lexical items. This (highly reliable) abstract knowledge is preferred over accrued experience of individual items, and it acts as gatekeeper to the processes of lexical activation and competition.

A Laboratory Phonology audience, as already noted, needs no reminder that lexical statistics are available to and important for language users, both in speaking and in listening. Of particular interest to this audience are gradient and word-specific effects, since they provide a modelling challenge for many traditional approaches in phonology. As described in the Introduction, exemplar theory has provided an attractive route for exploring alternative modelling solutions. In exemplar theory, individual processing episodes are veridically recorded and stored. As pointed out by Jusczyk (1997) and other modellers since, the early stages of speech acquisition in infancy virtually require episodic storage capacity. Infants are not explicitly taught sounds or words to start with; they only have what they hear, and what they hear is not isolated phonemes or even, usually, isolated words, but is mostly continuous speech (Van de Weijer, 1999; Aslin, Woodward, LaMendola and Bever, 1996). Yet before the end of the first year of life, infants have knowledge of what phoneme contrasts have relevance for the vocabulary of the environmental language (Werker and Tees, 1984), have knowledge of the typical lexical patterns of the environmental vocabulary (Jusczyk, Cutler and Redanz, 1993), and have begun to recognise words (Swingley, 2005). This can only have been achieved by drawing on the stored input, which, initially, has no abstract reflection at all.

Episodic representations are also required in modelling the adult lexicon to account for the abundant evidence of talker effects in speech perception (see, e.g., Johnson and Mullennix, 1997). But these effects are, like all the other effects discussed in the present paper, very selective, appearing where they are more useful; in word recall but not in lexical decision (Luce and Lyons, 1998), in shallow but not in deeper processing (Goldinger, 1996), in atypical but not in typical instances (Nygaard, Burt and Queen, 2000), and in differentiating words from nonwords only when the latter were word-like (*bacov*; McLennan and Luce, 2005). Thus the challenge facing speech perception modelling at the moment is accounting simultaneously for effects of abstraction and for effects of episodic storage (see Pisoni and Levy, 2007, and Goldinger, 2007, for attempts at a hybrid model). Although exemplar models allow the generation of abstractions (e.g., Wedel, 2007), the online use of abstract knowledge not arising online from the relevant distribution, as discussed in this paper, requires such a true hybrid model. (Moreover, pure exemplar models certainly cannot account for the perceptual retuning effects discussed here; Cutler et al., 2010).

The message to be extracted from the evidence reviewed here is that the way listeners chiefly draw on the statistics of their accumulated lexical experience is by converting them to abstract generalisations. This is the way they can best be deployed to facilitate future listening. Just as with the cases of category flexibility cited in the introduction to this paper, listening to speech is in this respect no different from other types of perceptual processing. Consider again the colours of traffic lights. In the European Union, the colours are standard, which means that the middle one is identical in Britain (where it is called *amber*), the Netherlands (where it is called *oranje* 'orange') and Germany (where it is called *gelb* 'yellow'). Ambiguous colours on a continuum from yellow to orange applied to a middle traffic light were categorised "yellow" more often by German-speaking viewers but "orange" more often by Dutch-speaking viewers, though this asymmetry did not appear when the same continuum was realised on socks (Mitterer, Horschig, Müsseler and Majid, 2009). Since the perceptual experience of traffic lights must have been the same for each group, the categorisations were driven not by perceptual but by declarative memory, which differed across the groups in that the middle traffic light was differently named. Abstract knowledge was thus drawn on to inform categorical interpretation where the input was ambiguous. Exactly the same happens in speech.

Correspondence e-mail address: anne.cutler@mpi.nl

References

Aslin, Richard N., Julide Z. Woodward, Nicholas P. LaMendola & Thomas G. Bever. 1996. Models of word segmentation in fluent maternal speech to infants. In James L. Morgan & Katherine Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* 117–134. Hillsdale, NJ: Erlbaum.

Beckman, Mary E. & Janet B. Pierrehumbert. 2003. Interpreting 'phonetic interpretation' over the lexicon. In John Local, Richard Ogden & Rosalind Temple (eds.), *Laboratory Phonology VI* 13–37. Cambridge: Cambridge University Press.

Bertelson, Paul, Jean Vroomen & Beatrice de Gelder. 2003. Visual recalibration of auditory speech identification: A McGurk after-effect. *Psychological Science* 14. 592–597.

Borsky, Susan, Betty Tuller & Lewis P. Shapiro. 1998. "How to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America* 103. 2670–2676.

Broadbent, Donald E., Peter Ladefoged & Walter Lawrence. 1956. Vowel sounds and perceptual constancy. *Nature* 178. 815–816.

Bybee, Joan L. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In Michael Barlow & Suzanne Kemmer (eds.), *Usage-based models of language* 65–85. Stanford: CSLI.

Connine, Cynthia M., Debra Titone, Thomas Deelman & Dawn Blasko. 1997. Similarity mapping in spoken word recognition. *Journal of Memory and Language* 37. 463–480.

Cutler, Anne & Sally Butterfield. 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language* 31. 218–236.

Cutler, Anne & David M. Carter. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2. 133–142.

Cutler, Anne, Frank Eisner, James M. McQueen & Dennis Norris. 2010. How abstract phonemic categories are necessary for coping with speaker-related variation. In Cécile Fougeron, Barbara Kühnert, Mariapaola d'Imperio and Nathalie Vallée (eds.), *Papers in Laboratory Phonology 10.* 91–111. Berlin: Mouton de Gruyter.

Cutler, Anne & Dennis Norris. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14. 113–121.

Cutler, Anne, James M. McQueen, Sally Butterfield & Dennis Norris. 2008. Prelexically-driven perceptual retuning of phoneme boundaries. In Janet Fletcher, Deborah Loakes, Roland Göcke, Denis Burnham & Michael Wagner (eds.), *Proceedings of INTERSPEECH, 9th Annual Conference of the International Speech Communication Association*, Brisbane, Australia, September 2008, 2056. Adelaide: Causal Productions (CD-ROM). ISSN 1990–9772. http://www.isca-speech.org/archive/interspeech_2008.

Dahan, Delphine, James S. Magnuson, Michael K. Tanenhaus & Ellen M. Hogan. 2001. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes* 16. 507–534.

Davis, Matthew H., William D. Marslen-Wilson & M. Gareth Gaskell. 2002. Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28. 218–244.

Eisner, Frank & James M. McQueen. 2005. The specificity of perceptual learning in speech processing. *Perception and Psychophysics* 67. 224–238.

Eisner, Frank & James M. McQueen. 2006. Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America* 119. 1950–1953.

Fougeron, Cécile, Barbara Kühnert, Mariapaola d'Imperio & Nathalie Vallée (eds.). 2010. *Papers in Laboratory Phonology* 10. Berlin: Mouton de Gruyter.

Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6. 110–125.

Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22. 1166–1183.

Goldinger, Stephen D. 2007. A complementary-systems approach to abstract and episodic speech perception. In Jürgen Trouvain & William J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*. 49–54. Dudweiler: Pirrot.

Hay, Jennifer, Janet B. Pierrehumbert & Mary E. Beckman, M. 2003. Speech perception, well-formedness, and the statistics of the lexicon. In John Local, Richard Ogden & Rosalind Temple (eds.), *Laboratory Phonology VI.* 58–74. Cambridge: Cambridge University Press.

Hay, Jennifer, Aaron Nolan & Katie Drager. 2006. From *fush* to *feesh*: Exemplar priming in speech perception. *The Linguistic Review* 23. 351–379.

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34. 458–484.

Hintzman, Douglas L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93. 411–428.

Jesse, Alexandra & James M. McQueen. submitted. Positional transfer in the lexical retuning of speech perception.

Johnson, Keith A. & John W. Mullennix. (eds.) 1997. *Talker variability in speech processing*. San Diego: Academic Press.

Jurafsky, Daniel, Alan Bell & Cynthia Girand. 2002. The role of the lemma in form variation. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7.* 3–34. Berlin/New York: Mouton de Gruyter.

Juscyk, Peter W. 1997. *The discovery of spoken language*. Cambridge, MA: The MIT Press.

Jusczyk, Peter W., Anne Cutler & Nancy J. Redanz. 1993. Infants' preference for the predominant stress patterns of English words. *Child Development* 64. 675–687.

Kraljic, Tanya, Susan E. Brennan & Arthur G. Samuel. 2008. Accommodating variation: Dialects, idiolects, and speech processing. *Cognition* 107. 51–81.

Kraljic, Tanya & Arthur G. Samuel. 2005. Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology* 51. 141–178.

Kraljic, Tanya & Arthur G. Samuel. 2006. Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review* 13. 262–268.

Kraljic, Tanya & Arthur G. Samuel. 2007. Perceptual adjustments to multiple speakers. *Journal of Memory & Language* 56. 1–15.

Kraljic, Tanya, Arthur G. Samuel & Susan E. Brennan. 2008. First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science* 19. 332–338.

Linden, Sabine van & Jean H. M. Vroomen. 2007. Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance* 33. 1483–1494.

Luce, Paul A. & Emily A. Lyons. 1998. Specificity of memory representations for spoken words. *Memory & Cognition* 26. 708–715.

Marslen-Wilson, William D. & Paul Warren 1994. Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review* 101. 653–675.

Maye, Jessica, Richard N. Aslin & Michael K. Tanenhaus. 2008. The Weckud Wetch of the Wast: Lexical adaptation to a novel accent. *Cognitive Science* 32. 543–562.

McLennan, Conor T. & Paul A. Luce. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31, 306–321.

McQueen, James M. 2007. Eight questions about spoken-word recognition. In M. Gareth Gaskell (ed.), *The Oxford handbook of psycholinguistics* (pp. 37–53). Oxford: Oxford University Press.

McQueen, James M., Anne Cutler & Dennis Norris. 2006. Phonological abstraction in the mental lexicon. *Cognitive Science* 30. 1113–1126.

Miller, Joanne L. & Alvin L. Liberman. 1979. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics* 25. 457–465.

Mitterer, Holger, Yiya Chen & Xiaolin Zhou. in press. Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science.*

Mitterer, Holger, Jörn M. Horschig, Jochen Müsseler & Asifa Majid. 2009. The influence of memory on perception: It's not what things look like, it's what you call them. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35. 1557–1562.

Mitterer, Holger & James M. McQueen. 2009. Foreign subtitles help but native-language subtitles harm foreign speech understanding. *PloS One* 4. e7785.

Mitterer, Holger & Jan Peter de Ruiter. 2008. Recalibrating color categories using world knowledge. *Psychological Science* 19. 629–634.

Morton, John & John Long. 1976. Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior* 15. 43–51.

Norris, Dennis, Sally Butterfield, James M. McQueen & Anne Cutler. 2006. Lexically-guided retuning of letter perception. *Quarterly Journal of Experimental Psychology* 59. 1505–1515.

Norris, Dennis, James M. McQueen & Anne Cutler. 1995. Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21. 1209–1228.

Norris, Dennis, James M. McQueen & Anne Cutler. 2000. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23. 299–325.

Norris, Dennis, James M. McQueen & Anne Cutler. 2003. Perceptual learning in speech. *Cognitive Psychology* 47. 204–238.

Nosofsky, Robert M. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115. 39–57.

Nygaard, Lynne C., S. Alexandra Burt & Jennifer S. Queen. 2000. Surface form typicality and asymmetries in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26. 1228–1244

Ohala, John J. 1995. The perceptual basis of some sound patterns. In Bruce Connell and Amalia Arvaniti (eds.), *Laboratory Phonology IV* (pp. 87–92). Cambridge: Cambridge University Press.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory Phonology 7* (pp. 101–140). Berlin/New York: Mouton de Gruyter.

Pisoni, David B. & Susannah V. Levi. 2007. Some observations on representations and representational specificity in speech perception and spoken word recognition. In M. Gareth Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 3–18). Oxford: Oxford University Press.

Pitt, Mark A. & James M. McQueen. 1998. Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language* 39. 347–370.

Salverda, Anne Pier, Delphine Dahan, Michael Tanenhaus, Katherine Crosswhite, Mikhail Masharov & Joyce McDonough. 2007. Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition* 105. 466–476.

Schreuder, Robert & R. Harald Baayen. 1994. Prefix stripping re-revisited. *Journal of Memory and Language* 33. 357–375.

Shatzman, Keren B. & James M. McQueen. 2006. Prosodic knowledge affects the recognition of newly-acquired words. *Psychological Science* 17. 372–377.

Sjerps, Matthias J. & James M. McQueen. 2010. The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 36. 195–211.

Swingley, Daniel. 2005. 11-month-olds' knowledge of how familiar words sound. *Developmental Science* 8. 432–443.

Vroomen, Jean & Beatrice de Gelder. 1995. Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 21. 98–108.

Vroomen, Jean, Monique van Zon & Beatrice de Gelder. 1996. Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition* 24. 744–755.

Weber, Andrea & Anne Cutler. 2006. First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America* 119. 597–607.

Wedel, Andrew. 2007. Feedback and regularity in the lexicon. *Phonology* 24. 147–85.

Weijer, Joost van de. 1999. *Language input for word discovery*. Ph.D. dissertation, MPI Series in Psycholinguistics, 9. University of Nijmegen.

Werker, Janet F. & Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7. 49–63.

Yaeger-Dror, Malcah. 1994. Phonetic evidence for sound change in Quebec French. In Patricia A. Keating (ed.), *Phonological Structure and Phonetic Form – Papers in Laboratory Phonology* III (pp. 267–292). Cambridge: Cambridge University Press.