The Culture-Phonology Interface

Commentary: The Role of Grammar in Adaptation and Imitation

Edward Flemming

MIT

web.mit.edu/~flemming/www

Perceptual adaptation and imitation

- Several of the studies presented here explore perceptual adaptation (Staum Casasanto) or imitation (Babel; Graff, Zuraw & Nielsen (GZN)) processes.
- There is evidence that adaptation and imitation are mediated by phonological grammar.
- To serve this function, phonological grammar must allow for variation (multiple outputs for a given input form).
- Production of variants must be systematic and socially conditioned, not random.

Imitation

- Listening to a speaker can result in a subject modifying their speech to be more similar to that speaker (Goldinger 1998 etc).
  - Progressive effect during shadowing (e.g. Babel).
  - Read word list, listen, read word list again (e.g. GZN).
- To interpret experiments based on this effect we need a model of this form of imitation.
- Hypothesis: Imitation (in this sense) is constrained by the speaker's grammar.
  - It is achieved by selecting among variants that are permitted by the speaker's grammar.
  - Where the grammar doesn't license appropriate variants, no imitation occurs.

A grammar-independent model of imitation

- Some previous accounts of imitation do not posit any role for grammar, e.g. Goldinger (1998).
- Episodic lexicon - each word is represented by many exemplars.
- Production target for a word is based on a weighted average of a subset of exemplars of that word (Pierrehumbert 2001)
- Imitation effect can be derived if there is a bias to select recent exemplars.

Imitation

- This model predicts that all properties should be imitated equally.
- If the bias to select recent exemplars is general, then imitation should always occur.

Selective Imitation

These predictions are incorrect:

- Not all properties are imitated equally:
  - Subjects imitate lengthened VOT but not shortened VOT (Nielsen 2007).
  - Subjects do not imitate /s/ realized as a sound between [s] and [ʃ] (Kraljic, Brennan & Samuels 2008).
  - Subjects imitate some vowels but not others (Babel)
- Degree of imitation depends on situation:
  - subject attitudes (Babel)
  - sex of subject/sex of speaker (GZN, Pardo 2006).

Selective Imitation

- Hypothesis: Imitation is achieved by selecting among variants that are permitted by the speaker's grammar.
  - Phonological grammars encompass variation - a given form has a range of well-formed realizations.
  - Imitation involves the selection of realizations from within this well-formed range that best correspond to the speech being imitated.

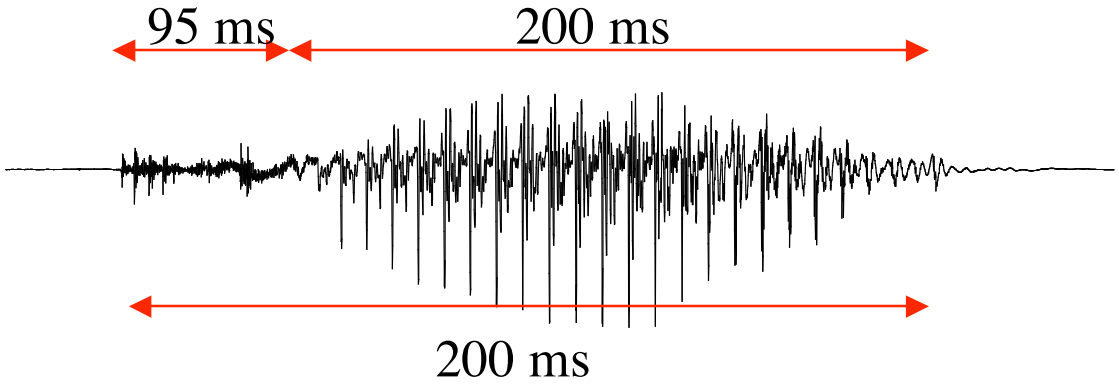
Imitation involves selection of grammatical variants

- Extended VOT is permissible, e.g. in hyperarticulation.
  - Smiljanic & Bradlow (2008): English word-initial VOT:
    - normal 78 ms, clear speech 120 ms.
- Reduced VOT is not permissible (except where required by increased speech rate).
  - Reduced VOT violates constraints on the distinctiveness of voicing contrasts (Nielsen 2007).

Illustrative phonetic grammar

VOT is determined by the interaction of three weighted constraints (Cf. Flemming 2001):

- MinDist(VOT): VOT of voiceless stop should be 90 ms greater than VOT of voiced stop of the same place of articulation (95 ms for [p^h], assuming [b] has VOT of 5ms)
  - cost of violation: $w_{VOT}(95\text{-}VOT)^2$
- Vowel length: Total vowel duration ($VOT+Vdur$) should be 200 ms.
  - cost of violation: $w_{dur}(200\text{-}(VOT+Vdur))^2$
- Fully Voiced Vowel: Voiced vowel duration should be 200 ms.
  - cost of violation: $w_{Vdur}(200\text{-}Vdur)^2$
- VOT is selected to minimize violation of these constraints (total cost)
  - $w_{\mathrm{VOT}}(80\text{-}VOT)^2 + w_{\mathrm{Vdur}}(200\text{-}Vdur)^2 + w_{dur}(200\text{-}(VOT+Vdur))^2$



Illustrative phonetic grammar

- VOT is selected to minimize violation of these constraints (total cost)
  - $w_{\mathrm{VOT}}(95\text{-}VOT)^2 + w_{\mathrm{Vdur}}(200\text{-}Vdur)^2 + w_{dur}(200\text{-}(VOT+Vdur))^2$
- If $w_{VOT} = 1$, $w_{Vdur} = 1$, $w_{dur} = 1$, then VOT = 63 ms
- If $w_{VOT} = 1.5$, $w_{Vdur} = 1$, $w_{dur} = 1$, then VOT = 71 ms
- Grammar can specify a range of possible weight values,
  - e.g. $1 < w_{VOT} < 1.5$
- Imitation can be achieved by selecting appropriate values within this range.
- Lengthened VOT is usually reserved for hyperarticulation, so baseline VOT is at the low end of the range.
- This leaves room for imitative lengthening, but not for shortening.

Imitation is constrained by grammar

- Analyzing imitation as grammar-based also provides the basis for an account of generalization in imitation.
- Subjects exposed to lengthened VOT in [p^h] without any examples of [k^h] lengthen VOT in both [p^h] and [k^h] (Nielsen 2007).
- Mindist(VOT) applies to all places of articulation, so selecting a weight for w_{VOT} affects VOT of all stops.

Imitation is constrained by grammar

- Subjects do not imitate /s/→[s/ʃ] (Kraljic et al 2008) - ambiguous [s/ʃ] is not part of the grammatical repertoire of most speakers.
  - NB mimicry can involve grammar modification.
- Grammar-based imitation may also play a role in explaining selective imitation of vowels in Babel's NZE study.

Selective imitation by NZE speakers

- NZE subjects imitated Australian /ɛ, æ/, but not / ɪ, ɑ, ʌ, ɔ/
- As Babel observes, /ɛ, æ/ are involved in a sound change in progress - both continue to raise (Trudgill et al 1998, Maclagan & Hay 2004).
  - implies significant synchronic variation in the heights of these vowels, so speakers are likely to command a range of variants of these vowels.
  - Since AusE /ɛ, æ/ are lower than their NZE counterparts, NZE speakers can use conservative variants to better approximate the AusE model.
- /ɑ, ʌ, ɔ/ are not undergoing comparable sound changes (and are less different from their AusE counterparts?)
- /ɪ/ centralized to [ə] in the 20thC.
  - This sound change is complete in these subjects?
  - The discrepancy from AusE [ɪ̟] is so large that a small shift in that direction would not yield a significantly improved match.

Imitation by CA speakers

- Only /æ, ɑ/ were subject to significant convergence but other vowels are subject to significant variation (e.g. /u/).
- Selective imitation for social reasons (Babel).
- Methodological point: these imitation studies lack a proper control condition.
  - Non-imitative factors could lead to change in the course of a shadowing task, or between two readings of a list of words. E.g. repetition might lead to reduction (cf. GZN).
  - To eliminate this possibility, it would be desirable to have a control group who listen to speech which is like their own in relevant respects.

Perceptual adaptation as grammar modeling

- Staum Casasanto (SC) examines perceptual adaptation rather than imitation.
- Hypothesis: Perceptual adaptation is also grammar based - listeners infer a speaker's phonetic/phonological grammar and use this model grammar to interpret speaker's productions (cf. Nielsen & Wilson 2008).

Perceptual adaptation as grammar modeling

- E.g. Kraljic & Samuels (2006)
  - Subjects listen to speech in which (a) realization of /t/ has been shifted towards /d/ (VOT lengthened etc) or (b) realization of /d/ has been shifted towards /t/ (VOT shortened etc).
  - Subjects categorize stimuli from /t-d/ and /p-b/ continua.
  - Perceptual boundary between voiced and voiceless stops differs between (a) and (b) groups - subjects shift the boundary in accord with the pattern of realization that they heard.
  - Shift generalizes to /p-b/ contrast although only /t-d/ was heard.
- This boundary shift can be analyzed as a result of the listener modeling the speaker's grammar for stop voicing and using the model to interpret speaker's productions.

Perceptual adaptation based on visual information

- SC shows that subjects perceptually adapt to the expected grammar of a speaker based on visual information.
  - they infer likely rate of t/d deletion from non-linguistic cues, i.e. visual information about the race of the speaker.
  - they use this information to interpret ambiguous utterances, e.g. [mæs] = *mass* or *mast*?
- people know something about the phonetic/phonological grammars of different social groups
- Where they can make inferences about group membership of a speaker based on non-linguistic information, they can guess likely properties of the speaker's grammar independent of any linguistic evidence.

Example

- E.g. Coetzee & Pater (2008) model frequencies of t/d deletion in different dialects using the following constraints:
  - *Ct: No clusters ending in a coronal stop
  - MaxC: An input consonant must have a correspondent in the output (‘don’t delete’)
  - MaxC/V: An input consonant preceding a vowel must have a correspondent in the output.
  - MaxC-Final: A phrase-final input consonant must have a correspondent in the output.
- Ranking these constraints derives different patterns of t/d deletion
  - deletion /_V, /_# → /_C

Perceptual adaptation as grammar modeling

- Coetzee & Pater (2008) show that rates of t/d deletion in different dialects can be analyzed as different rankings of these constraints in a Stochastic OT grammar (Boersma 1998).

	Constraint ranking values				Rates of t/d deletion by context		
Dialect	*CT	MAXC/_V	MAXC-FINAL	MAXC	/_V	/_#	/_C
AAVE (DC)[1]	101.0	102.3	96.8	99.0	0.29	0.73	0.76
Philadelphia[2]	107.2	108.2	110.6	92.8	0.38	0.12	1.00

[1]Fasold (1972), [2]Guy(1980)

  - Evaluation noise = 2

- In these terms, listeners would have some knowledge of characteristic constraint rankings for different groups.
- Analyzing this form of perceptual adaptation in terms of a model of the speaker's grammar provides a straightforward account of the fact that inferences about rate of t/d deletion apply equally to words and non-words.
  - inference concerns constraint ranking values, not words.

The nature of phonological grammars

- The analyses of imitation and adaptation outlined here presume that:
  - phonological/phonetic grammars encompass variation - i.e. they can generate multiple outputs for a given input.
    - E.g. /p/ can be realized with a range of different VOT values, /mæst/ can be realized as [mæs] or [mæst], etc.
  - variation can be indexed to particular social groups (etc).
  - variants can be selected according to context - e.g. to imitate speech patterns.
- Several current models of phonology allow for variation, but most treat variation as random - they provide no mechanism for external conditioning.
  - but see Hayes (2000), Coetzee & Pater (2008) for proposals to model careful-casual register variation.
- We have hypothesized that imitation involves the selection of constraint rankings/weightings from the range permitted by the grammar in order to better match another speaker's output.

References

- Boersma, Paul. 1998. *Functional Phonology. Formalizing the interactions between articulatory and perceptual drives*. Doctoral dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.
- Coetzee, Andries W., & Joe Pater (2008). The place of variation in phonological theory. To appear in J. Goldsmith, J. Riggle & A. Yu (eds.) *The Handbook of Phonological Theory, 2nd Edition.*
- Flemming, Edward (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18(1).
- Goldinger, Stephen (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251-279.
- Hayes, Bruce (2000). Gradient wellformedness in Optimalty Theory. Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer, eds., (2000) *Optimality Theory: Phonology, Syntax, and Acquisition*, Oxford University Press, pp. 88-120.
- Maclagan, Margaret, and Jen Hay (2004). The rise and rise of New Zealand English DRESS. Proceedings of the 10th International Conference on Speech Science and Technology, 183-188.
- Nielsen, Kuniko (2007). The Interaction between Spontaneous Imitation and LinguisticKnowledge. *UCLA Working Papers in Phonetics 105*. pp. 125-137.
- Nielsen, Kuniko, & Colin Wilson (2008). A Bayesian Network Model of Multi-level Phonetic Imitation. Proceedings of WCCFL 27.
- Kraljic, T., Brennan, S.E., & Samuel, A.G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. Cognition, 107(1), 51-81.

References

- Kraljic, T. & Samuel, A.G. (2006). Generalization in perceptual learning for speech. Psychonomic Bulletin and Review, 13(2), 262-268.
- Pardo, Jennifer (2006). On phonetic convergence during conversational interaction. *JASA* 119, 2382-2393.
- Pierrehumbert, J. (2001) Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee and P. Hopper (eds.) Frequency effects and the emergence of lexical structure. John Benjamins, Amsterdam. 137-157.
- Smiljanic, Rajka, & Bradlow, Ann (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics* 36, 91-113.
- Trudgill, Peter, Gordon, Elizabeth & Lewis, Gillian (1998). New dialect formation and Southern Hemisphere English: The New Zealand short front vowels. *Journal of Sociolinguistics* 2, 35-51.