Orthographic lengthening and phonological variation on Twitter
Gretchen McCulloch (Lingthusiasm) and Jeffrey Lamontagne (McGill University)

Introduction. Non-standard spelling in social media can reflect socially salient phonological processes (e.g. Tatman 2015, 2016). In English, people across dialects lengthen words in writing to add emphasis, which is done by repeating letters (ex. "way" > "wayyyy"). Our goal is to determine whether phonetic and phonological characteristics of different dialects predicts lengthening patterns – a phenomenon that, unlike word respellings (ex. "beah" for "beer"), does not obviously relate to the phonological variation or phonemic representations directly.

Previous work found that English lengthening typically targets word-final letters even if the letter is silent, with a weak preference for lengthening vowels rather than consonants (Lamontagne and McCulloch 2017). This study introduces dialectal and cross-linguistic components, using French data from three regions: Southern France, Northern France and Canada. In particular, we examine three cases: (a) final consonants, often silent; (b) final "e", which is sometimes silent and sometimes pronounced as schwa, at rates that vary by region (consistently in Southern France, rarely in Canada; Milne 2014); and (c) digraphs representing monophthongs, which don't map onto phonemes identically in the different varieties.

Methods. After collecting a corpus of over 65 million geotagged French tweets, we extracted 66 000 words in which at least one letter was repeated three or more times. We then performed mixed-effects logistic regression to predict which letters were most likely to be lengthened.

Results and discussion. Overall, we find a general preference to lengthen the final letter across all three varieties of French, like in English. This preference is still found when the last letter is a consonant that wouldn't be pronounced. Unlike in English, however, lengthening seems to target vowels as a primary goal, and will often either affect only the vowel or will affect all letters (silent or not) from the last vowel onwards.

Final "e" is nearly categorically lengthened when present, but it is the only letter lengthened about half as often as other vowels. It is sometimes respelled as "euh" to emphasize that the schwa is pronounced, in which case the "euh" is categorically targeted by lengthening. In southern France there's a greater probability of also lengthening the vowel of the previous syllable in polysyllabic words, reflecting that the schwa realisation patterns are different in this dialect. Lengthening "e" is more common beside consonants, paralleling consonant clusters' effects on schwa realisation in speech (Côté 2000; Milne 2014).

Finally, for digraphs representing monophthongs, we find that digraphs don't lengthen equally often: "ou" is less likely to lengthen than "au" and "eu", but the probabilities track whether the spelling maps onto a tense or a lax vowel in the specific dialect (suggesting of prosodic effects, *citation redacted*), suggesting phonemic effects are present even for identical spellings. Lengthening only "ou" is affected by the presence of a pronounced consonant in the Quebec data, which likely reflects the dialect's laxing process (Dumas 1987).

In summation, lengthening is similar in English and French in frequently targeting the final letter and showing letter-specific tendencies, but French lengthening targets the vowel as a priority. We also observe regional differences in speech inform variation in lengthening, which appear to reflect phonemic, phonological and prosodic patterns in the dialect. These results show that non-standard orthography can provide information about patterns in speech even when those patterns may not be salient or that are not directly encoded by the spelling.

References

Côté, Marie-Hélène. 2000. *Consonant cluster phonotactics: A perceptual approach.* PhD dissertation, MIT.

Dumas, D. 1987. *Nos façons de parler : les prononciations du français québécois.* Sillery : Presses de l'Université du Québec.

Gess, R., C. Lyche and T. Meisenburg. 2012. "Introduction to phonological variation in French". In: *Phonological Variation in French: Illustrations from three continents*. Philadelphia: John Benjamins. Pp. 1-22.

Milne, P. 2014. *The variable pronunciations of word-final consonant clusters in a force aligned corpus of spoken French.* Doctoral thesis submitted to the University of Ottawa.

Tatman, R. 2016. "'I'm a spawts guay': Comparing the Use of Sociophonetic Variables in Speech and Twitter". *Selected Papers from NWAV 44 (22: 2):* 161-170.

Tatman, R. 2015. #go awn: Sociophonetic variation in variant spellings on Twitter. In Working Papers of the Linguistics Circle 25.2: Proceedings of the 31st annual NorthWest Linguistics Conference, ed. S. Onosson and M. Huijsmans, 97–108.

Werry, C. M. 1996. "Linguistic and interactional features of Internet Relay Chat". In S. C. Herring, ed., Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives: 47-63.