Morpheme structure constraints

Phonotactic theory (EGG 2025, Zagreb)

Close reading

- Halle 1978:95:

> The native speaker of a language knows a great deal about his language that he was never taught. An example of this untaught knowledge is illustrated in (1), where I have listed a number of words chosen from different languages, including English. In order to make this a fair test, the English words in the list are words that are unlikely to be familiar to the general public, including most crossword-puzzle fans:
>
> (1) ptak thole hlad plast sram mgla vlas flitch dnom rtut
>
> If one were to ask which of the ten words in this list are to be found in the unabridged Webster's, it is likely that readers of these lines would guess that *thole*, *plast*, and *flitch* are English words, whereas the rest are not English. This evidently gives rise to the question: How does a reader who has never seen any of the words on the list know that some are English and others are not? The answer is that the words judged not English have letter sequences not found in English. This implies that in learning the words of English the normal speaker acquires knowledge about the structure of the words. The curious thing about this knowledge is that it is acquired although it is never taught, for English-speaking parents do not normally draw their children's attention to the fact that consonant sequences that begin English words are subject to certain restrictions that exclude words such as *ptak*, *sram*, and *rtut*, but allow *thole*, *flitch*, and *plast*. Nonetheless, in the absence of any overt teaching, speakers somehow acquire this knowledge.

- **Why does Halle say this knowledge is "untaught"?**
- **Is this knowledge language-universal or language-specific?**

1 Introduction

- Phonotactics (or *phonotaxis*) is the subfield of phonology concerned with generalizations about permissible sound sequences in languages.
- Descriptively, there are constraints on possible sound sequences in nearly every language. But what—if any—role do these constraints play in grammar?

- We can subdivide this big question into three smaller ones:
  - To what degree do speakers encode phonotactic generalizations?
  - Do speakers encode phonotactic generalizations at the level of underlying representations (as *morpheme structure constraints*), surface representations (as *surface structure constraints*), or both?
  - Are speakers' phonotactic generalizations *autonomous* (i.e., distinct) from their (narrow) phonological knowledge or *projected* (i.e., derived) from it?

2 Morpheme structure constraints

- Chomsky and Halle (1965), henceforth C&H, claim that speakers encode redundancies in underlying representations, like the following for English:

(2) $\left[+\textsc{Consonantal}\right] \rightarrow \left\{+\textsc{Liquid}\right\} / \# \begin{bmatrix}+\textsc{Consonantal}\\ -\textsc{Continuant}\end{bmatrix} \_\_ \left[+\textsc{Vocalic}\right]$

- In prose, this rule says that any consonant after a word-initial stop and followed by a vowel becomes a liquid.
- C&H posit these redundancy rules because they are committed to the belief that the lexicon is maximally underspecified, with all major redundancies handled by early rule.
- They assume the /ɹ/ in the underlying form of *brick* is not specified +Liquid; rather, (2) inserts that specification at the start of the derivation.[1]

(3) Underlying representation of *brick* (after C&H):

$$\left\{\begin{array}{c}+\textsc{Labial}\\ -\textsc{Coronal}\\ -\textsc{Dorsal}\\ +\textsc{Consonantal}\\ -\textsc{Continuant}\\ -\textsc{Sonorant}\\ -\textsc{Liquid}\\ +\textsc{Voice}\\ \dots\end{array}\right\}\left\{\begin{array}{c}-\textsc{Labial}\\ +\textsc{Coronal}\\ -\textsc{Dorsal}\\ +\textsc{Consonantal}\\ +\textsc{Continuant}\\ +\textsc{Sonorant}\\ \emptyset\textsc{Liquid}\\ \emptyset\textsc{Voice}\\ \dots\end{array}\right\}\dots$$

- Stanley (1967) discerns two types of *morpheme structure constraints* (MSCs):
  - *segment structure rules* fill in redundancies in the underlying phoneme inventory.
  - *sequence structure rules* fill in redundancies in the phoneme sequences in URs.
- C&H claim (2) predicts that

[1] In his Presidential Address to the Linguistics Society of America, Halle (1975) seems to reject this approach, attributing it to a youthful, overenthusiastic abuse of information theory.

  - *brick* is an actual word of English and *blick* is a possible word, but
  - *bnick* is not a possible word.

- How exactly does (2) predict those facts?

(4) Redundancy Rule Projection Principle: Let A, C, D be natural classes and B a phonological change. Then, if there exists a redundancy rule $A \rightarrow B \,/\, C \_\_ D$, any word containing the substring CAD is not a possible word.

- It is not clear to me if (4) is valid as written: a projected constraint on URs need not be *surface-true*.
- For example, voicing is non-contrastive for Russian affricates *ц, ч* /t͡s, t͡ɕ/ (Halle 1959:22). The following segment structure constraint expresses this generalization.

(5) $\left[+\text{DelRel}\right] \rightarrow \left\{-\text{Voice}\right\}$

- However, these affricates have voiced allophones [d͡z, d͡ʑ] on the surface (e.g., [ʐe**t͡ɕ**] 'to burn vs. [ʐe**d͡ʑ** bɨ] 'would burn').
- These sort of observation leads Clayton (1976) and Shibatani (1973), among others, to reject MSCs in favor of *surface structure constraints*, to which we turn tomorrow.

3 A case for nihilism?

- In an oft-cited passage, Hale and Reiss (2008), henceforth H&R, write:

> Ridding phonological theory of MSCs is clearly a laudable goal, since such constraints serve merely to state descriptive generalizations about the memorized content of the lexicon of a particular language. Even if we, as linguists, find some generalizations in our description of the lexicon, there is no reason to posit these generalizations as part of the speaker's knowledge of their language, since they are computationally inert and thus irrelevant to the input–output mappings that the grammar is responsible for. Now, one might argue that this is what all phonological statements achieve: a generalization over a body of individual tokens. For example, the aspiration rule of English just restates a generalization about aspiration in surface forms of English. However, the difference between MSCs and phonological rules should be clear: the former generalize over stored, i.e. memorized, information, whereas the latter are meant to capture productive patterns of generated information. (H&R:17–18)

- H&R claim not all regularities/tendencies in the lexicon (i.e., in the inventory of segments and sequences in URs) are encoded by the speaker (call this H_0). In contrast, others believe speakers encode any significant generalization.

(6) H_1: Let g be a statable phonotactic generalization. Then g is active in the grammar if it is statistically robust (in a sense to be defined).

> …the patterns outlined above are statistically significant. Given this, it stands that these sound patterns should be explained by some linguistic mechanism. (Brown 2010:48)

- H&R further claim that phonotactic generalizations are "computationally inert", in contrast to phonological generalizations which are computationally active. **Why do H&R make this contrast?**
- In Finnish, for example, all native words end in either a vowel, a coronal consonant (e.g., /s, t, n, r, l/), or a glottal stop. **What is the status of this fact in the minds of Finnish speakers? What would H&R say about this?**
- Some questions remain:
  - How can we discern which regularities speakers encode or ignore?
  - How would H&R account for the wordlikeness judgments in (1)?

    > Imagine that you were to ask a naïve non-linguist monolingual English speaker to discern whether a short snippet of spoken language was either, say, Māori or Croatian. Let us suppose that this speaker does not know a single word of either language. Would this speaker do better than chance at this task? (Charles Reiss, p.c.)

- Oh et al. (2020, 2023) report that monolingual English-speaking New Zealanders are able to discriminate between "possible" and "impossible" Māori nonce words.

References

Brown, Jason. 2010. *Gitksan phonotactics.* LINCOM.

Chomsky, Noam, and Morris Halle. 1965. Some controversial questions in phonological theory. *Journal of Linguistics* 1:97–138.

Clayton, Mary L. 1976. The redundance of underlying morpheme-structure conditions. *Language* 52:295–313.

Hale, Mark, and Charles Reiss. 2008. *The Phonological Enterprise.* Oxford University Press.

Halle, Morris. 1959. *Sound Pattern of Russian.* Mouton.

Halle, Morris. 1975. Confessio grammatici. *Language* 51:525–535.

Halle, Morris. 1978. Knowledge unlearned and untaught: What speakers know about the sounds of their language. In *Linguistic Theory and Psychological Reality*, ed. Morris Halle, Joan Bresnan, and George A. Miller, 294–303. MIT Press.

Oh, Yoon Mi, Simon Todd, Clay Beckner, Jen Hay, and Jeanette King. 2023. Assessing the size of non-Māori-speakers' active Māori lexicon. *PLOS ONE* 18:e0289669.

Oh, Yoon Mi, Simon Todd, Clay Beckner, Jen Hay, and Jeremy Needle. 2020. Non-Māori-speaking New Zealanders have a Māori proto-lexicon. *Scientific Reports* 10:22318.

Shibatani, Masayoshi. 1973. The role of surface phonetic constraints in generative phonology. *Language* 49:87–106.

Stanley, Richard. 1967. Redundancy rules in phonology. *Language* 43:393–436.