Phonological grammar is traditionally stipulated to involve the manipulation of categorical representations by general operations that at worst have only a few lexical exceptions. The fact that lenition is sensitive to lexical frequency calls these stipulations into question. A review of the evidence, including a recent study on frequency effects in syllable contraction in Southern Min, shows that a gradient increase in lexical frequency is associated with a gradient decrease in the duration, temporal separation, and/or spatial extent of articulatory gestures (as estimated by their acoustic correlates). The gradience is not an illusion caused by overlapping categorical processes, frequency doesn't merely affect the probability of choosing among alternative categorical representations, and the phenomenon is not reducible to contextual predictability.

These observations are consistent with the exemplar model of speech production, where the production lexicon consists of clouds of phonetically detailed exemplars, but they do not fit well within phonological grammar as currently conceived. All of the current proposals to handle frequency effects within grammar fail on technical grounds, and they only get as far as they do by ignoring the facts that both frequency and lenition are gradient. When gradience is taken into account, the best formal analysis seems to be to assign a separate cophonology to every item in the lexicon, a conclusion consistent with the exemplar model. Moreover, grammatical universals are expected to show up in first-language acquisition, but higher frequency causes greater lenition only in adult speech. Children show a reversed pattern of an apparently different nature, a "practice makes perfect" effect that is readily handled by existing learning models. This suggests that the adult pattern is not learned. This is again consistent with the exemplar model, where lenition is caused by production.

Taking frequency effects in lenition seriously has two major phonological implications. First, not all phonologically relevant types of patterns can be accommodated within (the same type of) grammar; a complete theory of phonology must be modular. Second, given that lenition is markedness reduction in the raw, and that it is represented in the mind grammar-externally, the possibility is raised that grammar is not in the markedness business at all, contrary to a fundamental claim of Optimality Theory.