Phonotactic restrictions on consonant vowel sequences are one of the most exotic features of Japanese segmental phonology. For example, while both [s] and [t] are regularly occurring sounds in spoken Japanese, their combination, that is [st], is absent from the standard dialect. These phonotactic gaps are not all equal. Diachronic sound changes, which took place between Early and Contemporary Japanese, reveal a delicate hierarchy among the CV restrictions. Some gaps are easier to fill than others. For example, [ti] sequences appeared relatively early, around the pre-war period (e.g., ‘party’ > [parti] Arakawa 1932: 99), while [tu] is considered to be an unstable formation even now (e.g., ‘two’ > [tsu:], ‘Bantu’ > [bantu]–[bantsu]). It can be argued that the diachronically unfolding pattern follows certain perceptual characteristics of the involved CV contrasts. The basic insight of the perceptual approach is that consonantal contrasts are sensitive to vocalic context. Vowels can greatly diminish important perceptual cues such as formant transitions, or release bursts. A syllable fails to emerge if its consonant cannot maintain a sufficiently salient perceptual contrast with other consonants in a given context (cf. P-map: Steriade 2001; dispersion: Padgett 2001). For example, the lack of /ho/-/fu/ contrast in Japanese is due to the slightly labial, but not rounded, [u] context that impedes the perception of contrastive labial cues (cf. Foulkes 1997). The same reason is behind the occasional neutralization of the otherwise existing /ho/-/fu/ contrast (e.g., ‘formant’ > [hurumanto], ‘smartphone’ > [sumaho]). While the perceptual approach can explain various implicational aspects of CV restrictions, some of the pervasive phonotactic gaps call for a more categorical explanation. The goal of this research is to investigate how mainstream representational approaches can account for these phonotactic gaps, and how the emergence of innovative syllables (i.e., their phonologization) can be explained from a representational point of view.

From the abundance of phonological representations (cf. Krämmer 2012), the current research singles out three representative ones: (1) a descriptivist approach, (2) a classical derivational approach, and (3) an underspecificational approach. The descriptive approach relies heavily on the concept of minimal pairs. If a minimal pair or an appropriate near minimal pair exists for two sounds in a language, then they are regarded as separate phonemes. According to this view /s/ and /ʃ/ are separate phonemes in Japanese (Bloch 1950: 113). The lack of [si] on the surface goes back to the lack of *si/ at the underlying level. Since the birth of innovative syllables are simply viewed as novel combinations of consonants and vowels, this approach does not add much to our understanding of Japanese phonotactics. The classical derivational approach, on the other hand, provides a mechanism that explains the absence of certain syllable types in Japanese. In the derivational paradigm, surface forms are derived through assimilatory rules. For example, [tʃi] is derived through a /tʃ/ ↔ /ʃ/ assimilation (Tsujimura 2007: 32). The assimilatory rules are present not only in lexical adaptations (e.g., ‘ticket’ > [tʃiketto]), but are also present in the native phonology (e.g., verb conjugations). The derivational approach correctly predicts that those syllables will be difficult to adapt whose forms coincide with the inputs of assimilatory rules (cf. Hattori 1960). This intuitively attractive explanation, however, has problems explaining the emergence of new CV forms. For example, an innovative [ti] cannot go back to /ti/ as /ti/ is subject to the above mentioned assimilation. There are several ways to circumvent the unwanted derivational step, but as the sheer variety of existing approaches imply (e.g., lexical domain-specific rules in McCawley 1968: 62-75 and Ih & Master 1995; fossilization in Hattori 1960), these theoretical adjuncts do not resolve the controversy convincingly. An intuitively more attractive explanation can be achieved by relying on underspecified representations. For instance, the [s] ↔ [ʃ] contrast can be analyzed as an opposition between a sibilant /ʃ/ and an /S/ phoneme, the latter of which is underspecified for place of articulation. The underspecified /S/ may receive either a default alveolar value (e.g., /Sa/ → *[sa]; see Lahiri & Reetz 2002 for default place), or borrows a palatal feature from a subsequent /i/ vowel (i.e., /Si/ → *[ʃi]). This analysis is on par with the derivational approach in explaining CV gaps, adaptation patterns and conjugation paradigms (e.g., /oS+i/ → *[ou], /oS+i/ → *[oʃ]). Its advantage over other approaches lies in its analysis of the phonologization process. The phonologization of the [si] ↔ [ʃi] contrast in an underspecificational framework is considered to be difficult because it requires the listener to develop a place-specified /s/ phoneme parallel to the already existing /S/ and /ʃ/. This new /s/ is not subject to place-feature spreading (/si/ → *[si]). Note that this explanation neither requires special features (e.g., *[foreign]), nor assumes special lexical domains in order to suspend otherwise acknowledged phonological processes such as spreading or assimilation. Phonologization is assumed to be a process that only involves the manipulation of underlying representations. Although the resulting tripartite consonant system of sibilants (i.e., /S/+ʃ/+/s/) and its likes (e.g., /T-/ʃ-/t/) might seem awkward, it describes a transitional stage from partially neutralizing contrast (/S/+ʃ/) to a state in which the opposition may occupy all vocalic contexts (/ʃ+/s/). Furthermore, interpreting phonologization as a representational development is supported by feature matching theories of speech perception (Lahiri & Reetz 2002), which in turn provides further interfaces to the perceptual explanations mentioned above.
Bibliography


