L2/24-049

THE INTERNATIONAL PHONETIC ASSOCIATION

Registered Office: INTERNATIONAL HOUSE, 24 HOLBORN VIADUCT, LONDON EC1A 2BN, UK

from The President

1st January 2024

Unicode support for historical and para-IPA letters

Introduction

The International Phonetic Association wishes to lodge a formal request for Unicode support of various historical and para-IPA letters and diacritics, including modifier versions of letters that are currently encoded. This document outlines the consultation process that took place prior to this request, explains the motivation for the proposal, and provides some details on the scope of the request.

This request follows on Unicode proposals L2/20-252 and L2/20-253, which covered modifier versions of nearly all modern IPA letters. Historical and para-IPA notation had been largely deferred until the Council of the International Phonetic Association and consulting phoneticians could more fully discuss the desired scope of Unicode support. The Alphabet, Charts and Fonts Committee of the IPA has now had a chance to review which symbols have adequate usage in recent literature, or form integral sets with such symbols, to warrant formal support. Requested symbols fall into three categories: superscript modifiers, per L2/20-252 and L2/20-253; affricate ligatures; and superscript alternatives to subscript IPA diacritics.

Consultation

Prompted by an initial approach from K. Miller, the President wrote an internal discussion document dealing with the issue, which was circulated to the members of the Alphabet, Charts and Fonts Committee in September 2023. Miller's proposal had been subdivided into related sets of symbols for individual consideration by the IPA. These were reviewed by the Alphabet, Charts and Fonts Committee, which has unanimously agreed to support the following:

- The affricate ligatures ⟨dð tθ dɮ tɬ tꞎ d𝼅⟩ and ⟨ʧ̡ ʤ̡⟩.
- Superscript alternatives to the IPA and extIPA diacritics ⟨◌̙ ◌̘ ◌̠ ◌̺ ◌̻ ◌̼ ◌͇ ◌̚ ◌͎ ◌͢◌⟩.
- The double caron ⟨◌̌̌⟩. We understand that this has been requested, based on its appearance in the Report of the Kiel Convention (IPA 1989), but that no decision has been made by the UTC.
- Modifier versions of:
  - the pre-Kiel click letters ⟨ʇ ʗ ʖ 𝼋 ψ ʞ⟩.
  - the voiceless implosive letters ⟨ƥ ƭ 𝼉 ƈ ƙ ʠ⟩.
  - the Sinological extensions to the IPA ⟨ȶ ȡ ȵ ȴ ᴀ ᴇ ω ɿ ʅ ʮ ʯ⟩.
  - the common palatal-hook letters ⟨ᶁ ŋ̡ ᶊ ᶎ⟩ (⟨ᶵ ᶪ⟩ are already encoded).
  - the historical and unofficial letters ⟨ᴅ ȸ ȹ ʤ̡ ʤ̢ ɥ̶ ɟ η ɼ ʆ ʆ̡ ʧ̡ ʧ̢ ɵ w̶ ʒ̡ ʓ⟩. In a few cases the baseline letters are also needed – namely barred ⟨w̶ ɥ̶⟩ and the voicing pair ⟨ʧ̡ ʤ̡⟩, which we understand are being separately proposed.

One concern raised by members of the Committee is that the rounded forms of the click letters U+0297 and U+1DF0B illustrated in K. Miller's proposal (⟨ʗ⟩ and ⟨𝼋⟩, and thus modifier ⟨ᶜ⟩ and ⟨ᶠ⟩) differ from the straight forms ⟨ʗ⟩ and ⟨𝼋⟩ preferred by the IPA. The worry of the Committee is that Unicode might present these rounded forms as prescriptive and that they would consequently be reproduced as the default forms in fonts. However, the glyphs displayed for U+0297 and 1DF0B in the published Unicode charts closely match the forms preferred by the IPA. If Unicode models the modifier glyphs after the current code charts, that will adequately address the concerns of the Committee. The present proposals can therefore be regarded as coming jointly and severally from the entire Committee of six members, who are listed below as individual signatories, as well as from myself as President of the IPA.

Motivation

In September 2020, the Council of the International Phonetic Association lodged a formal request for general Unicode support of IPA letters used as superscript modifiers, noting that "the consensus of the IPA Council is that a codepoint should be allocated for the superscript version of every letter-like symbol used in IPA notation, and indeed that a Unicode implementation which lacks this is incomplete." The requested characters were accepted by the UTC with proposals L2/20-252 and L2/20-253. The 2020 request was largely restricted to the letters of the current IPA alphabet, and we noted at the time that "a more methodical approach to historical and para-IPA symbols may be formulated after further consideration by the Council." Since then, the *Alphabet, Charts and Fonts Committee* was established to evaluate the encoding requirements of the IPA alphabet, preferred glyph forms, and similar matters, and it is this committee that has reviewed the current proposals.

The reason that a formal request is needed from the IPA is that the desired set of symbols is only partially attested in a survey of the literature. Restricting Unicode coverage to those attested symbols, while omitting accidental gaps, would limit the utility of the symbols that are encoded. In addition, if only the subset that have hitherto been documented were to be included, that would bias Unicode toward particular languages and phonological models. For example, Sinological ⟨ⁿɖ̥⟩ is attested in K. Miller's sources, but not ⟨ɳ̊ᵈ⟩, which might be equally expected; the Sinological vowels ⟨ᶦ ᶤ⟩ are attested, but not their rounded equivalents ⟨ᶣ ᶹ⟩, presumably because baseline ⟨ɥ̟ ɥ̠⟩ are used for far fewer languages than are ⟨ɭ̟ ɭ̠⟩. Similarly, the voiceless historical affricate ⟨t͡ʃ͡ˠ⟩ is attested in IPA publications but not its voiced partner ⟨d͡ʒ͡ˠ⟩, which can be expected to be less common. Before the adoption of L2/20-252 and L2/20-253, members of the IPA Council had encountered difficulties with online presentation of material transcribed in the current IPA alphabet due to such gaps in support, and similar difficulties can be expected from accidental gaps in para-IPA and historical IPA coverage, in the digitization of historical documents and for authors who continue to use such symbols.

Scope of the proposal

We request greater Unicode support for three categories of IPA letter variant:

- superscript modifier letters,
- affricate ligatures, and
- superscript alternatives to subscript diacritics.

Superscript modifiers: Superscripting is a crucial dimension of IPA representation that requires hard encoding for preservation of the underlying data structure. In the opinion of the IPA Council, reported in Ashby (2020), modifier versions of "historical letter-like symbols should be supported if [the historical symbols] are attested in the recent literature." As was done in L2/20-252 and 253, accidental gaps in the literature need to be filled, as it would be detrimental to the practical use of the alphabet if Unicode were to formalize such gaps. In the unanimous opinion of the Alphabet, Charts and Fonts Committee, the set of letters listed in the "Consultation" section above meets these criteria.

We do not request modifier support for all historical IPA letters. Indeed, several historical letters are not supported by Unicode even in their baseline forms, as they are obsolete and there is no apparent need for them at present. Examples include ⟨ꜧ⟩, the graphic ancestor of ⟨ɤ⟩; ⟨ɬ̟⟩, the voiceless partner to ⟨ɮ̟⟩; ⟨ʓ⟩, the voiced partner to ⟨ʒ⟩; 'hissing-hushing' ⟨ʗ⟩; and proposed but never officially adopted letters such as ⟨ɲ⟩. Archaic letters that have Unicode support as baseline characters, but for which we do not expect to need modifier versions, include ⟨ɋ ʁ ʀ ɧ ꞯ ɪ⟩ for modern ⟨ʕ χ ʁ̥ ɬ̠ ɸ ɪ⟩ and 'hissing-hushing' ⟨ʓ⟩. More recent letters that we do not see sufficient need to encode as

modifiers include the labialized (whistled) sibilants ⟨ʃʷ ʒʷ ɬʷ ɮʷ⟩, the affricates ⟨ʦ ʣ⟩, r-coloured vowel letters such as ⟨ɚ⟩, and the sporadic letter variant ⟨ɞ⟩. Some of these might be requested in the future if evidence surfaces of their recent use.

Affricate ligatures: At present we only see a need for affricate ligatures of the coronal (tongue-tip) consonants of the IPA proper. Non-coronal affricates such as ⟨kx gɣ⟩ and para-IPA coronal affricates such as ⟨tʂ dʑ⟩ are not requested.

Superscript diacritics: We request superscript alternatives to those modern diacritics that are not yet supported. Among historical diacritics, we understand that a superscript turned omega, ⟨◌̒⟩, is being requested separately based on attestation in the IPA literature. We find no need for a superscript alternative to the 'open' diacritic ⟨◌̩⟩, although one might be needed for related phonetic notation systems. The rare superscript variant of the palatal hook ⟨◌̡⟩ also has no apparent need at present.

We do not address Americanist notation or typewriter substitutions of IPA letters, apart from the overlap in modifier ⟨ᴅ⟩ and coincidentally with ⟨ʷ⟩. Although the Council feels it is important for Americanist and typewriter notation to be well supported by Unicode, as already noted in Ashby (2020), that is too broad a topic to cover here.

References

Michael Ashby, 23 September 2020. Unicode support for IPA letters as superscript modifiers. Letter submitted to the Unicode Technical Committee.

International Phonetic Association, 1989. Report on the 1989 Kiel Convention. *Journal of the International Phonetic Association*. 19 (2): 67–80.

Signed on behalf of the IPA Council,

Katerina Nicolaidis (President) Aikaterini Nikolaidou Aikaterini Nikolaidou 05.01.2024 14:31

The members of the Alphabet, Charts and Fonts Committee for 2023–2027 are:

Prof. Dr Adrian P Simpson	Germany
Dr. Michael Ashby	U.K.
Prof. John Esling	Canada
Prof. Asher Laufer	Israel
Prof. Lee Wai Sum	Hong Kong
Benjamin V. Tucker	U.S.A.