

**DoNotEmit for Arabic tashkil**

Roozbeh Pournader

2024-12-18

This is a proposal for adding two lines to DoNotEmit.txt for basic Arabic tashkil as well as corresponding text for the Core Spec.

The author was going through Microsoft's OpenType specification for Arabic script and what it calls "Invalid Combining Marks" (see <https://learn.microsoft.com/en-us/typography/script-development/arabic#handling-invalid-combining-marks>). He noticed that while he disagrees with that specification's approach of disallowing sequences that indeed occasionally occur in Quranic Arabic as well as some minority orthographies in the Arabic script (see UAX #53, Section 5.5), there are some things that can be borrowed from that specification.

Specifically, no one considers a sequence of two fathas or two kasras valid, and indeed they can be used as a mechanism to confuse a user or system. They also have valid replacement characters. (The same cannot be necessarily said for two dammas, since a dammatan is often visually distinct from two dammas.)

Thus, the author proposes that the following two lines be added to DoNotEmit.txt:

```
064E 064E; 064B; Arabic_Tashkil # ARABIC FATHA, ARABIC FATHA; ARABIC FATHATAN  
0650 0650; 064D; Arabic_Tashkil # ARABIC KASRA, ARABIC KASRA; ARABIC KASRATAN
```

Proposed text for the Core Spec (probably for the section *Tashkil Nonspacing Marks*) follows:

- Note that many tashkil combinations may be considered "impossible" in the regular Arabic language, but indeed occur in Quranic Arabic or other languages written in the Arabic script. But specifically, due to visual confusability, using a pair of fathas and a pair of kasras is discouraged in favor of fathatan and kasratan.