

Left is Right, Right is Left: Joining_Type for Left-to-right Scripts

Author: Roozbeh Pournader

Date: 2026-05-13

For consideration by the Unicode Script Encoding Working Group and the Unicode Technical Committee

Summary: This document provides some background about the `Joining_Type` property and discusses options for the `Joining_Type` of joining left-to-right scripts.

Unicode has recently received two proposals for *joining* left-to-right scripts, which would benefit from so-called “Arabic” joining properties.

One of the scripts, N’ti, is made of all dual-joining letters, which should receive `Joining_Type=D`. The other, Ndiko Jonam, has some dual-joining letters, as well as some other letters that join to their right side only. This raises the question of which `Joining_Type` they should get, R (short for `Right_Joining`) or L (short for `Left_Joining`). This is a new question because there are no joining left-to-right scripts in Unicode yet. Almost all joining scripts are right-to-left, with only a couple of top-to-bottom ones (Mongolian and Phags-pa).

The header of the file `ArabicShaping.txt` currently says:

“Note that for cursive joining scripts which are typically rendered top-to-bottom, rather than right-to-left, `Joining_Type=L` conventionally refers to bottom joining, and `Joining_Type=R` conventionally refers to top joining. See Section 14.4, Phags-pa for more information on the interpretation of joining types in vertical layout.”

Section 14.4 further says:

“Because the Phags-pa script is typically rendered from top to bottom, `Joining_Type = L` (`Left_Joining`) conventionally refers to bottom joining that is, joining to a character which follows (is below) it. `Joining_Type = R` (`Right_Joining`) is not used for the Phags-pa script, but would refer to top joining, that is, joining to a character which precedes (is above) it. Most Phags-pa characters are `Dual_Joining`, as they may join on both top and bottom.”

This is all good and well, except that both Mongolian and Phags-pa are rendered left-to-right quite frequently as a fallback in environments which don't support top-to-bottom display. In such environments, which are very common, you can already see left-to-right joining scripts in practice.

The Arabic Cursive Joining Algorithm, if we may call it that, is described in Section 9.2.2 of the Core Specification. After a brief reference to the Bidirectional Algorithm, the cursive joining algorithm is defined in detail in terms of visual order. The only exception is that it says:

“An implementation may choose to restate the following rules according to logical order so as to apply them *before* the Bidirectional Algorithm’s reordering phase. In this case, the words *right* and *left* as used in this section would become *preceding* and *following*.” (Generally, I believe it would be confusing to try to redescribe the whole algorithm in terms of preceding and following as opposed to left and right.)

Going back to Mongolian and Phags-Pa, it should be noted that all their letters are dual-joining, except for a single character: U+A872 PHAGS-PA SUPERFIXED LETTER RA, which has Joining_Type=L (Left_Joining). According to the above information, it means that the letter is “bottom-joining”.

Now, in a left-to-right display of Phags-pa, U+A872 is indeed right-joining visually. It does not join to the letter to its left, while it does join to the letter to its right. This is visible in sequences such as <ka, superfixed ra, ka, ka> which render as follows:



Figure 1. Renderings of a made-up Phags-pa word

Since Unicode is considering encoding joining left-to-right scripts, we have two options going forward:

- 1) Left is Right, Right is Left: Codify the existing practice for U+A872 and for future Ndiko Jonam characters, define the visually right-joining characters in left-to-right scripts as L (Left_Joining). This is the most backwards-compatible option, but could become confusing for users and implementers of the standard when we actually try to specify it. In this system, we would need to say that right really means preceding, and left really means succeeding. We may even want to define some new property aliases to clarify L and R.

- 2) Left is Left, Right is Right: Change the `Joining_Type` of U+A872 from L (`Left_Joining`) to R (`Right_Joining`), and encode the future Ndiko Jonam right-joining characters as R (`Right_Joining`). This has the benefit of a simple specification, but would require some existing Phags-pa implementations to change how they interpret `Joining_Type`. It would also require an explanation about the change for top-to-bottom scripts in the Phags-pa section of the Core Spec.

It appears to the author that the cleanest way forward for the specification is the second option (Left is Left, Right is Right).

A quick check on the specification of Universal Shaping Engine did not result in any changes necessary there, although some implementations of Phags-pa driven by Unicode joining properties may need to be updated.