

The visual and typographic dimension of language: analysis and clustering of languages, based on language design features

Marta Guidotti | University of Hasselt, Belgium

The visual rhythm of text, often perceived as a ‘black and white stripe pattern’, is fundamentally shaped by a language's typographic design features, including letter frequencies, combinations, word lengths, and capitalization. Building on Unger's [1] principles, these inherent textual rhythms give a solid base when reading and are crucial for visual comfort, impacting spatial frequency [2].

However, prevalent legibility research, predominantly focused on English, overlooks this critical linguistic diversity within Latin script languages. Despite designers recognizing these visual nuances, current layout software lacks the functionality to address language-specific typographic needs. This widespread gap highlights the urgent need for a method supporting feasible cross-language legibility research and design by exploring how these language determined visual rhythms affect legibility.

This research addresses this challenge by developing a novel system to categorize and group Latin script languages based on their shared typographic and visual features, while also identifying the most decisive features for each group. We meticulously examined 72 languages, analyzing 34 distinct typographic design features that define variations in text patterns. For example, the frequency of diacritics [a], ascenders [b] and descenders [q] change the blackness in the space between the text lines, while long or short words define the proportion of black elements to white spaces in the text line. Through comprehensive frequency analysis of multilingual corpora and hierarchical clustering, we successfully identified four distinct language groups. We can see, within a theoretical framework and data-visualisation, how these interplay with established language-families, and the historical, geographical and geopolitical background of languages.

This novel, data-driven visual approach to language offers an unprecedented framework for conducting truly cross-language legibility research. Furthermore, it provides designers with a powerful tool, encouraging conscious consideration and integration of language-specific visual characteristics into layout practices, ultimately fostering more inclusive, perceptually rich, and comfortable reading experiences worldwide.

Unger, Gerard 2006. *Il Gioco della Lettura*. Viterbo: Nuovi Equilibri. (Terwijl je leest, 2006)

Bessemans, Ann 2025. The automatization of rhythm deduction in type [Automatic Type Design 3]

Wilkins, Arnold J. 1995. *Visual Stress*. Oxford University Press: London