Dimitrios Meletis, Jonas Romstadt

The world's punctuation systems: Comparing forms, functions, and variation

Abstract

Punctuation is found in most writing systems despite their formal and typological diversity (consider, for example, morphographic Chinese and alphabetic English). The reason for its universality is not only cross-system borrowing but also, and more importantly, its cognitive and communicative relevance. This is the starting point of our paper: First, we dissect previous linguistic conceptualizations of punctuation, focusing on its distinction from so-called 'default graphemes' (such as alphabetic letters or morphographic characters). Then, by analyzing a sample of punctuation inventories from 30 diverse writing systems, we show that, on the one hand, there exist striking formal and—even more so—functional similarities between individual punctuation units as well as entire punctuation inventories. We suggest that this is because punctuation serves the universal function of facilitating reading and comprehension processes. On the other hand, we also find unique formal and functional features and variation because punctuation must also meet other criteria (such as fitting language-specific structures or reinforcing sociolinguistic independence). By additionally drawing on historical and psycholinguistic data, we provide a first comparison of the world's punctuation systems and discuss how it can be a basis for comparative punctuation research.

1. Introduction

Despite the gradual establishment of an interdisciplinary study of writing—which we call grapholinguistics—research on punctuation has remained sparse. The few exceptions, which include Geoffrey Nunberg's *The Linguistics of Punctuation* (Nunberg 1990) or the work of German linguist Ursula Bredel (see below), are focused on specific languages (in this case English and German, respectively). By contrast, comparative punctuation research that "includes language systems, sociolinguistic norms, and pragmatic practice is still a desideratum"[1]. Aside from punctuation often being treated as an aside in typological research on writing systems, what hinders a comparison of different systems of punctuation—and writing systems in general (Meletis 2021)—are language barriers. Indeed, if the punctuation of/in a language *is* treated at all, it is likely in prescriptively oriented grammars or style guides published in that language, making them inaccessible to many. However, this often constitutes no great loss as these treatments often remain superficial, focusing on norms rather than describing the actual usage of punctuation.

Ironically, punctuation units themselves transcend such language barriers regarding their forms and especially their functions. One obvious reason for this is that countless writing

[1] Taken from the Call for Papers for the Conference *Comparative Punctuation Worldwide*, on which this volume is based.

systems have merely adopted a pre-existing 'Western'[2] system of punctuation that originated in ancient Greece and Rome and was rounded out in (post-)Medieval Europe; its most prominent units are the comma and the full stop. However, the fact that there also exist unique systems of punctuation exhibiting formal and functional similarities with this widespread 'Western' set points to a greater universality due to another, more relevant factor: the cognitive and communicative relevance of punctuation.

Due to limited space, this paper can only provide a first step in a comprehensive comparison of the world's punctuation systems by addressing the following questions: How do punctuation units distinguish themselves from intuitively 'normal' graphemes and what does this mean for their linguistic status? What are commonalities and differences of the punctuation units in some of the world's diverse writing systems—and why do they exist? To investigate these questions, we will present several observations regarding punctuation (Section 2) before discussing previous attempts at assessing it linguistically (Section 3). We will then explore formal and functional commonalities and differences as well as their possible motivations (Section 4), after which we draw conclusions and give an outlook for future comparative punctuation research (Section 5).

2. Relevant aspects for the comparative study of punctuation

Punctuation is constant. In writing systems, punctuation complements so-called default graphemes such as 'letters' in alphabets or 'characters' in Chinese (Meletis 2020), which relate linguistic units (phonemes, syllables, morphemes) with the shapes of given scripts (e. g. Roman or Cyrillic script). What is striking is that punctuation is far more universal than scripts. Indeed, a small set of 'Western' punctuation units[3] including the comma and the full stop is rather consistent. As our survey of the punctuation inventories of 30 diverse writing systems shows (see Section 4), of 28 that use a script other than Roman, 12 use the identical inventory that is used in German and English (i. e., ‹ . ; , : - – … ’ ? ! ( ) “” ›[4]), while 12 use a subset of it;[5] only 4 use a completely different system. This constancy of punctuation concerns both its forms and functions, although the latter appear to be more universal as they remain the same even if the specific forms may vary (cf. the full stop in writing systems using Roman script ‹.› vs. in the Chinese and Japanese writing systems ‹ 。 ›). One extralinguistic reason for this observed universality is, of course, pervasive borrowing (cf. Section 4.1). An 'internal' reason could be the role that punctuation assumes in cognitive processing.

---

[2] We realize that the broad term 'Western' is not unproblematic, which is why we use it in single quotation marks. One inspiration for using this umbrella term is Parkes' (1993) influential *Pause and effect: Punctuation in the West*.

[3] Note that terminologically, we speak of punctuation *units* here because the two predominantly used terms carry connotations that may have theoretical implications: 'mark' focuses on the materiality of punctuation units while 'sign' is tied to semiotics and insinuates that, like default graphemes, punctuation units have a value (in the sense of a signified), which is one of the questions discussed in this paper.

[4] This inventory for German and English is abstracted from graphetic variation of individual punctuation marks. For example, quotation marks are commonly realized like this in German: ‹„“› (although variation can also be observed here, cf. Fuhrhop, Reinken & Romstadt 2023). In English they normally have the form ‹“”›. Nevertheless, these punctuation *marks* correspond to the same punctuation *unit* in the given inventory.

[5] This means that at least half of their punctuation units stem from this 'Western' inventory; in 4 instances, only one unit is different—in Bassa, Cree Plains, Western Kayah and Manding (see Appendix).

In the context of comparative grapholinguistics, it bears to keep in mind that writing systems (including punctuation) must meet different requirements simultaneously, with an analysis of how well they do so proving useful for multi-perspective comparisons (cf. Meletis 2020): Most importantly, writing systems must structurally suit their languages (**linguistic fit**), be fitting for physiological and cognitive processing by human writers and readers (**processing fit**), and be accepted socioculturally and politically by/in literate communities (**sociocultural fit**). As we will argue below, different scripts and the default graphemes they materialize are governed mostly by the linguistic fit, whereas punctuation units cater predominantly to the processing fit. In other words, default graphemes are made to fit languages, punctuation units users—and first and foremost readers.

Punctuation relates to many linguistic levels. A given writing system's default graphemes are typically in a (primary)[6] relation with one linguistic sublevel (a notable exception being Japanese with its multiple subsystems). The graphemes in alphabets, for example, relate to phonemes, the graphemes in the few contemporary morphographic writing systems (such as Chinese) to morphemes. This relation is at the core of writing system typology, underlying the assumption of different types (cf. Joyce & Borgwaldt 2011).[7] Punctuation behaves differently, which has been one of the main challenges in deciding whether punctuation units count as graphemes. In general, they do not homogeneously relate to specific linguistic units but serve as aids for the processing and comprehension of written utterances at diverse levels. Rather than to mark specific linguistic structures, they reflect cognitive operations during production or perception processes. 'Diverse' is the keyword: punctuation units relate to different levels including the morphological, syntactic, textual, prosodic, and pragmatic levels, affecting different stages of language processing. This functional versatility of punctuation has rendered attempts of integrating all punctuation units into a single coherent framework elusive, calling for two shifts in our treatment of punctuation: on the one hand, from a static and purely descriptive linguistic approach to a more dynamic, process-oriented view (cf. Bredel 2008). On the other hand, from viewing punctuation units as constituting a neat 'set' to accepting them as formally and functionally heterogeneous and incongruent among themselves.[8]

Punctuation underlines the (relative) independence of writing and its idiosyncratic features. Lastly, even though punctuation has often been linked to speech and, like the written modality in general, reduced to being a depiction of what we 'hear' (cf. Baudusch 1980), the manifold relations with different linguistic levels and a strong association with the spatiality of writing (see below) underline that punctuation is one of the written features that call for an independent and interdisciplinary study of writing. Notably, this does not contradict that punctuation units "can likewise be considered as having a suprasegmental parallel in speech" (Henderson 1985: 144). However, writing and reading differ crucially from speaking and hearing. Here, punctuation may aid in uncovering some of the core differences—among them

[6] A secondary level must also be considered as graphemes may convey information beyond the relation to a specific type of linguistic unit: Chinese graphemes, for instance, relate primarily to morphemes but include phonographic components that relate to sound.

[7] Ideas for classifying writing systems (also) on the basis of punctuation are addressed by Bartsch (1998: 517).

[8] This, of course, does not mean that there are not meaningful subsets in which units share formal and/or functional commonalities (cf. also Bredel 2008).

cognitive ones. This positions comparative punctuation research at (or very near) the core of a burgeoning interdisciplinary field known as grapholinguistics.

3. Previous linguistic assessments of punctuation

It is widely accepted that punctuation (just like word spacing) developed for the needs of silent reading. The 'punctuational big bang' (Houston 2014) is thought to have occurred in ancient Greece. This term highlights how the introduction of punctuation and spacing marked a revolutionary development as the processing load of capturing and comprehending the structure of texts was shifted from reading aloud to reading silently. Phylogenetically, punctuation is a 'secondary system' as it developed long after the primary default graphemes. Related to this is the interpretation (and/or marginalization) of punctuation as a 'surplus', i. e. an (optional) add-on to writing systems.

Defining what constitutes a punctuation unit is by no means trivial. Unsurprisingly, thus, in early studies, the criteria that underlie the classification of different classes of units of writing systems often remained implicit. This is different in Bredel's (2008, 2020) analysis of the autonomous segmental means of the German writing system (see Tab. 1);[9] her model can also be adapted for other—at least alphabetic—writing systems.

Tab. 1: Bredel's classification of different graphic resources

	letters	**digits**	**special characters**	**punctuation units**	**blank spaces**
identifiable without context	+	+	+	+	–
recodable	+	+	+	–	–
freely combinable	+	+	–	–	–
paired	+	–	–	–	–

Bredel differentiates between letters, digits, special characters, punctuation units, and blank spaces. Note that she only focuses on segmental means of writing and therefore does not take into consideration subsegmental elements (such as diacritics) or higher-order structures. Relevant for us are the distinctions between punctuation units and other graphically overt segmental means on the one hand as well as blank spaces on the other. The feature of identifiability captures that punctuation units have a graphic form and are thus independently 'visible', i. e., without their graphic surroundings, whereas blank spaces are 'invisible' and depend on non-blank material surrounding them. Recodability acknowledges that default graphemes such as letters or special characters are pronounceable given that they have a phonological representation—such as ‹&› 'and'—whereas punctuation units are not. Following these criteria, punctuation units are identifiable but not recodable.[10] Against this background, we can ask how they relate and functionally compare to the other units in a writing system. Here, differences in the recodability of the different categories of units already give a first indication.

[9] For an earlier, though not as formalized approach, see Catach (1980).

[10] Note that in several other conceptions, such distinctions are not made, and the blank space is treated as a punctuation unit (cf. Jones 1995); Houston (2013: XI) calls it "invisible punctuation".

This question of the function of written elements leads almost inevitably to the concept of *grapheme* (at least in structuralist works). Like the definition of the grapheme overall (cf. Meletis 2019), the inquiry into whether punctuation units qualify as graphemes, or more broadly, how to classifly them graphematically, has been marked by a complex history. Indeed, in most studies, the status of punctuation units remains unclear. As an example, Baudusch (1981: 206)[11] refers to punctuation (in German) as "a special graphic subsystem" in which punctuation units "behave similarly to graphemes" (Baudusch 1981: 207). This vagueness is symptomatic of many graphematic publications.

Puncuation is more explicitly treated by Gallmann (1985), who generally considers graphemes "the smallest structural units of written language that can be defined both formally and functionally" (Gallmann 1985: 10). For him, punctuation units are, formally, independent "graphemes in the narrower sense" (Gallmann 1985: 11) closely related to other segmental means of writing. Functionally, he conceptualizes them as so-called syngraphemes (from Greek *syn-* 'with, together, same'; cf. Gallmann 1996; already used in Veith 1985). In other words, punctuation units co-occur with other graphemes. This distinguishes them from these other graphemes, the default graphemes ('Grundgrapheme', Gallmann 1985: 18).

Günther (1988: 77), by contrast, defines graphemes by analogy with phonemes as "the smallest meaning-distinguishing units of the writing system of a language"; this does not straightforwardly apply to punctuation units (but see below for counterexamples). Recently, Berg (2019: 29) has further specified this, positing that "graphemes are the smallest distinctive, syllabically autonomous units of the written language". This excludes most punctuation units as they cannot—on their own—occupy positions within the graphematic syllable.[12] What remains unclear, however, is how punctuation units operating at the word level such as apostrophe, hyphen, and abbreviation period (cf. Buchmann 2015 for these units in German) behave in this respect.[13]

All previous remarks refer to discussions of the grapheme that focused primarily on the writing system of German (and, in exceptions, English). By contrast, Meletis (2019) proposes a comparative concept of grapheme applicable also to other types of writing systems: "The grapheme can be defined as a basic unit of writing that (1) distinguishes meaning, (2) has a linguistic value (typically by referring to a linguistic unit), and (3) is minimal in that it is not composed by smaller units which are themselves graphemes" (Meletis 2019: 43). In asking whether punctuation units meet these criteria, we can, firstly, observe that they can sometimes distinguish meanings; however, this distinction is made not at the lexical but rather at the syntactic and/or pragmatic levels (cf. *Hello!* vs. *Hello?*), which can be regarded a secondary effect (cf. Gallmann 1985: 27). Secondly, the linguistic value of punctuation units is rather abstract, which is closely connected to their non-recodability and the fact that they interact with many linguistic levels (see Section 2). More recently, Meletis & Dürscheid (2022: 142) have argued that punctuation units "do not correspond with linguistic units of any kind". Their conclusion is that if punctuation units are to be "modelled as graphemes, the definition of

[11] We translated all German quotes in this paper ourselves.

[12] For the concept of the graphematic syllable, see, in German, Fuhrhop & Buchmann (2009), or, in English, Evertz (2018: Chapter 3).

[13] Particularly abbreviations including periods often appear like well-formed graphematic syllables, e. g., ‹esp.› for 'especially'.

grapheme has to be extended (which makes it more imprecise) or another class of graphemes has to be defined that behaves differently both structurally and functionally".[14] We are leaning towards the second option.

This question of the linguistic value of punctuation units is related to their general function. Already more than 140 years ago, Alexander Bieling (1880: 5) claimed that they "make the infinite series of letters [...] clearer for the eye by grouping" them, which is in accord with Gallmann (1985: 27) classifying them as 'comfort graphemes'. This position is best pronounced in Bredel's (2008, 2020) perception-oriented 'online' perspective that views punctuation units as reading aids. Specifically, Bredel (2008: 102) argues that a description of punctuation units can consist of "specifying the language processing mechanism underlying each individual unit" in the form of "x instructs the reader to do y" (Bredel 2020: 5), with x being a punctuation unit (e. g. full stop) and y being a task in language processing (e. g. stop syntactic parsing). As for the cognitive reality of this claim, psycholinguistic studies on different writing systems have shown that readers utilize punctuation in language processing (see Section 4.1). The fact that the general function of punctuation is tied to reading necessitates a re-evaluation of what 'graphem(at)ic' can mean, as this concept has thus far been used almost exclusively in the context of static—in the words of Bredel, 'offline'—description that focuses on how punctuation marks linguistic structures.

Regarding the difference between default graphemes vs. punctuation units, it can be posited that the former's main function is to convey meaning as they are combined to form larger structures such as graphematic words, sentences, or even entire texts. Thus, while a segmental analysis reveals that they have individual values, the primary level at which meaning is constituted is suprasegmental—most prominently that of written words (cf. Stetter 1997; Schmidt 2018). These words have meanings and, as building blocks of parsing processes, serve as important structures in the (first) perceptual and cognitive processing of written texts. This marks a major functional difference between default graphemes and punctuation units that is captured by the feature of combinability: default graphemes as well as digits can combine with each other to constitute (larger) meaning-bearing units, whereas punctuation units are not freely combinable; their graphotactics are constrained as they mostly attach to default graphemes, digits, or special characters.

Considering these differences, Bredel's classes can be subsumed under two categories: content graphemes by analogy with content words, which include default graphemes and digits, and function graphemes (or maybe more fittingly 'structure graphemes') by analogy with function words, which include punctuation units and blank spaces (cf. Bredel 2020: 8).[15]

What do these assessments of punctuation mean for our study? Three main aspects can be condensed: (1) Punctuation units differ from default graphemes both functionally and formally. In some respects, they behave like default graphemes, in other respects they differ. (2) Treating punctuation units as if they formed a formally and functionally coherent set may be part of the problem rather than the solution. For example, punctuation units at the word level behave differently from sentence-level units. (3) Nonetheless, punctuation units fulfil a common basic function: They serve as reading aids that support diverse cognitive operations.

---

[14] In French, Catach (1980: 21) called this latter class *ponctème*; English *puncteme* is found in, e. g., Krahn (2014: 199).

[15] Special characters lie somewhere in the middle as they have a content but are not freely combinable.

4. Comparative survey

4.1 Method and challenges

In this study, we compare 30 punctuation inventories. The writing systems they belong to are diverse regarding (a) the languages they represent (i. e., the language families these can be categorized into), (b) the types of writing systems they belong to (alphabets, abjads, abugidas, syllabaries, or morphographic systems; cf. Meletis & Dürscheid 2022: Chapter 6), as well as (c) the scripts that they use, with 28 of them employing a script other than Roman. A table with all systems is provided in the Appendix. The data was taken mainly from r12a,[16] a comprehensive resource for information on different scripts that itself was collected from various sources. Note that, as emphasized in r12a, the exact uses of given punctuation units in given writing systems may be contended as not enough primary data or reliable secondary sources are available (yet), which reflects the perhaps greatest challenge of comparative punctuation research that we already mentioned.

Against this background, our survey cannot aim to provide fine-grained analyses of individual systems but rather paints a first comparative picture. Notably, our results are skewed towards a formal analysis as we can more readily assess the forms of punctuation units in a writing system than their precise functions. Regarding the latter, the comparative basis of our study is the 'Western' set of punctuation units that is also at the core of Bredel's (2008, 2020) analysis. Thus, if the r12a page for a writing system provides an information like Balinese ‹᭟› being "[e]quivalent to a comma"[17], then it is assumed that its function is roughly similar to that of the 'Western' comma, and it is placed in the respective column of our comparison matrix. This, of course, means that our function-related conclusions must be regarded as preliminary as more research is necessary to assess the exact language-specific functions.

Another limitation of our comparison is that it is exclusively synchronic, which diminishes its explanatory force. For example, as we established above, many writing systems use (a subset of) 'Western' punctuation regardless of the languages they represent or the scripts that they use. Since we often lack the information of how a given punctuation inventory developed or how different inventories influenced each other, we can only hypothesize about the reason for this striking universality. A relatively safe assumption that has been confirmed for several writing

[16] https://r12a.github.io/scripts/ (March 14, 2024).
[17] https://r12a.github.io/scripts/bali/ban.html#word (March 14, 2024).

systems[18] is that punctuation was often borrowed from one system to another.[19] Indeed, this applies not only to punctuation but to writing systems in general, as most of them derive from few original and independent creations, seemingly reducing the explanatory power of synchronic cross-linguistic commonalities (cf. Sampson 2016). However, we argue that it is not only the independent emergence of a phenomenon that points to its particular significance but likewise the degree to which it has been adopted and, even more importantly, prevailed for such a long time and in diverse systems—as is the case for punctuation. Indeed, both independent emergence and adaptation/perseverance call for an explanation. In other words, while borrowing processes may be heavily influenced by sociolinguistic and language political factors (e. g., language contact or language ideologies[20]), in the case of punctuation, their stability is likely linked to their cognitive and communicative relevance.

In the following, we will make general assumptions regarding the functions and use of punctuation units (Section 4.2) before addressing commonalities and differences in their forms (Section 4.3).

4.2 Functions

4.2.1 The online view

Due to limited space as well as a lack of literature on the specific functions of punctuation units across diverse writing systems, we will focus on four of the most frequent units in our data: the 'syntactic' units full stop (occurring in all 30 systems) and comma (29/30), and the 'pragmatic' units parentheses (26/30) and quotation marks (24/30). In Bredel's (2020) analysis of the

[18] Four relatively well-documented examples are: (1) In Arabic, punctuation was introduced by newspaper editor and professor ʾAḥmad Zakī after he had travelled to France and found French written language easier to read than Arabic (see also Awad 2015 for forerunners of Zakī). (2) In Chinese, which traditionally exhibited some types of punctuation-like units (including correction marks, repetition marks, abbreviation marks, but also segmentation marks, which "[come] closest to our modern notion of punctuation" (Galambos 2014: 350; see also Richter 2023), "in April 1919, a group of scholars, including Hu Shi 姆祝 (1891–1962), presented a 'Proposal for the Implementation of New-Style Punctuation' to the first meeting of the Preparatory Committee for the Unification of the National Language" (Hamm 2020: 136). A revised version of this proposal became official in 1920, including twelve punctuation marks: the period, comma, semicolon, colon, question mark, exclamation mark, quotation mark, em dash, ellipsis mark, parentheses, proper name mark, and title mark (Mullaney 2017; Hamm 2020). Since 2011, 17 marks are officially used (Stryjewska 2016). (3) Japanese, while also being influenced by Chinese, was affected by European punctuation practices; Twine (1984: 233) describes how the Dutch studies scholar Furukawa Masao was the first to use full stops and commas in a primary reader for education in 1872, and identifies the Meiji period as an important time for experimentation with 'Western' punctuation, especially in Japanese translations from European languages. (4) Finally, Lee (2014) suggests that Korean adopted 'Western' punctuation in the late 19th century through Japan.

[19] Other possible factors influencing similarities in punctuation inventories are belonging to the same language family and regional proximity (consider, for example, the formal and functional commonalities of Balinese, Buginese, and Javanese, all Austronesian languages used in Indonesia).

[20] While an utterly interesting topic, (language) ideologies associated with punctuation cannot be addressed in detail here. However, it is noteworthy that the devaluation of a lack of punctuation (or the existence of only 'rudimentary' punctuation) appears to be a recurring topos. Consider two examples: For Arabic, Meynet (1971: 88, cit. from Awad 2015: 124) argues that "[a]ccepting to reform Arabic writing means accepting to change the bases, and to recognize that this instrument [meaning the Arabic language] is now inadequate, defective, and that the other [Europe], once again, is right." For Chinese, Bodde (1991: 15) writes that "a crude proto-punctuation can be found in certain Chinese texts going back as early as the third century B.C., yet until recent times punctuation in China remained under-developed, sporadically used, and often looked down upon as a vulgar device unworthy of the true scholar." Evidently, ('Western') punctuation is often treated as something desirable and distinguishes developed from 'inadequate' or 'crude' and 'vulgar' systems.

punctuation system of German, the full stop and the comma concern sentence-level processing and the cognitive dimension (i. e., parsing) while the parentheses and quotation marks serve text-level processing and indicate to readers shifts in communicative roles (reader vs. writer). Their specific functions are collected in Tab. *2*.

Tab. 2: A processing-oriented view of four frequent punctuation units (Bredel 2020)

punctuation unit	reading instruction
full stop	instructs readers to complete syntactic parsing and pass the structure that is marked by the full stop on to textual parsing
comma	instructs readers not to parse the structures on the left and right of the comma as subordinating and subordinated unit (i. e., part of the same syntactic constituent); it marks coordination, dislocation, or a sentence-internal syntactic boundary (cf. Primus 2007)
parentheses	indicates to readers additional information that does not change the truth value and/or grammaticality of the sourrounding structure (cf. Ström Herold/Levin 2021)
quotation marks	indicates to readers that the writer rejects responsibility/authorship for the passage enclosed by quotation marks (either because it is quoted or to express reservation of content and/or style) (cf. Klockow 1980)

This online view with its focus on processing-oriented functions is, of course, just one side of the coin: punctuation units do also mark syntactic or textual structures at which these cognitive operations must then be performed. And while it is likely that these operations (if broady defined) remain relatively stable across languages, the linguistic specifics can vary; for example, whether a comma is required at (restrictive) sentence-internal phrase boundaries.[21] In other words, methodologically, the online view may fittingly capture the main function of punctuation, which is arguably universal, but in comparative punctuation research, it must be accompanied by and synchronized with offline analyses of the actual linguistic implementation of specific punctuation units in given writing systems. Indeed, as Lindbüchl (2014: 83) found in her comparison of the comma in German, English, and French, offline constructions affect online procedures, and only the comma in German comprehensively provides the reading instruction of marking disruptions in syntactic parsing. This also calls for a re-evaluation of certain (possibly Eurocentric) assumptions, such as that punctuation is only indirectly related to prosody through its connection to syntax or semantics (cf. Primus 2007), which may not be the case for, e. g., the comma in French and other writing systems in our sample.

We argue that what is also needed in a holistic approach is actual psycholinguistic evidence of whether punctuation units actually perform these theoretically assumed functions.[22] Most of the published studies focus on the comma in a single language (mostly English), leading to a 'doubled' research gap in a field that is as a whole marginalized: On the one hand, there is a lack of research on the actual effect of other punctuation units. Select studies show that the full stop is often associated with a **sentence wrap-up effect**, which is analyzed in the context of

[21] For example, in English, it is ‹Dogs that bark don't bite›, whereas in German it is ‹Hunde, die bellen, beißen nicht›.

[22] This is not something Bredel (2020: VII) was concerned with: "It is not about concrete reading behavior but about structuring requirements during reading, which can be theoretically deduced from the principles of language processing."

language acquisition by Tiffin-Richards & Schroeder (2018). To name another example, Schlechtweg (2022) finds that advanced readers of English comprehend so-called pure quotations as in *"Gold" is a concrete noun* more quickly if quotation marks are used. Schlechtweg & Härtl (2023) observe an analogous effect for the detection of irony when indicated by quotation marks.

On the other hand, we also know little about the punctuation units in other writing systems, making cross-linguistic psycholinguistic studies a crucial desideratum (few do exist, e. g., for the comma in Japanese in ambiguous sentences, cf. Niikuni & Muramoto 2014). Thus, we can theorize about the function of punctuation units, which can also be of merit, but what they actually *do* still needs to be investigated further. Such studies could help us understand, for example, why the units in Table 2 are the most frequent ones, but also why others occur rarely[23] and why some punctuation inventories are smaller than others.

4.2.2 Specific cross-linguistic functions

In this section, we address some of the types of units that we find across writing systems that are not necessarily present in the 'Western' set. They are already found in older punctuation systems (such as medieval Chinese, cf. Galambos 2014), implying that they also serve important universal functions.

A point that should be emphasized upfront is that some of the units listed as examples here may not be in regular use today; reliable information about usage is challenging to gather and assess given the lack of literature. However, even if given units are not used (anymore), their mere existence is also of importance.

Repetition. Several units are used to indicate iteration. In an online view, they give the reading instruction 'repeat the preceding value'—whether it is that of a syllable, word, phrase, etc.[24] Examples include Thai ‹ๆ›, indicating that a word should be repeated and Cambodian ‹ៗ› for the repetition of a word or phrase. Javanese has ‹ꦃ› to mark that a syllable should be repeated, while Buginese has the formally identical ‹ᨃ› to indicate the doubling of a root or word. Japanese also has iteration units for kanjis ‹ 々 › as well as hiragana ‹ ゝ › and katakana ‹ ヽ › graphemes. In Indonesian, Filipino, and Malay the Arabic numeral ‹2› has previously served this purpose. Historically, Epigraphic Mayan has a duplication 'diacritic', "two dots optionally and rarely affixed to another grapheme to command the reader, in the majority of cases, to read a syllabogram twice in sequence" (Mora-Marín 2024: 1). In medieval Chinese manuscripts, optional devices indicate the repetition of a single character or a string of identical characters (cf. Galambos 2014: 345f.). As for a classification, the main question regarding these units is whether they can be categorized as punctuation: in an offline view, they receive a value (mostly) from the preceding unit, so they are pronouncable; however, this value is a contextually dependent one.

Abbreviation. Another type of unit that recurs indicates abbreviations. In many writing systems—such as English, German, and Russian—the period serves double duty as a full stop

[23] The ones occurring in the fewest systems are the apostrophe (5/30) and the dash (10/30).

[24] In German and English the ellipsis ‹…› can be used to indicate repetition, but that is not the only function of that punctuation unit and is, at least in German, a rather marginal one (cf. Romstadt accepted).

and an abbreviation marker.[25] This is also the case in Thai, in which it has been "borrowed from the West" and where another unit is used as an elision mark, with the reader having to supply the omitted information: ‹ฯ› (Haas 1980: 86f.). In Hindi, ‹°› serves as an abbreviation marker, so that, for example, ‹रुपया› 'rupee' can be abbreviated as ‹रु°›. In sinistrograde Hebrew, where the period is also sometimes used to mark abbreviations (but is regarded nonstandard), the *geresh*, which looks like a single quotation mark or apostrophe, is the standard unit for this purpose: ‹גברת› is abbreviated as ‹גב׳›.[26] The formal similarity to the 'Western' apostrophe is noteworthy as the apostrophe also marks omitted material that needs to be reconstructed by readers (cf. Bredel 2020: 40–43). Another unit used for abbreviations and sometimes classified as a punctuation unit (cf. Nunberg, Briscoe & Huddleston 2002: 1726) is the slash ‹/›: in English, it is used in two-letter initialisms such as in ‹w/o› for *without*, and in Modern Greek it is sometimes used in abbreviations such as ‹α/φοί› for ‹αδελφοί› 'brothers' (example taken from r12a). In sinistrograde Assyrian Neo-Aramaic, a suprasegmental horizontal line above a string of graphemes and with a circle at the beginning, the middle, and the end indicates abbreviations (cf. Fig. 1).

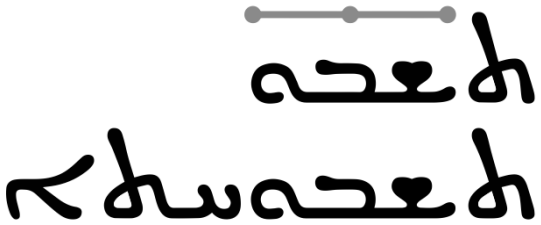

Fig. 1: Abbreviation (above) and full word in Assyrian Neo-Aramic, from: https://r12a.github.io/scripts/syrc/aii.html#abbrev

Lists. For in-line lists (i. e., coordinations), which are marked by the 'Western' comma in many writing systems, there are also other options, such as the Chinese enumeration comma *dùnhào* ‹、› that exists beside the 'normal' comma (*dòuhào*) (cf. Stryjewska 2016). In Japanese and Korean, the interpunct ‹·› assumes this function (as well as hyphen- and dash-like functions). Of course, there are also other ways to indicate lists, which involve special spatial arrangements or what has been called the *list mode* (vs. the *text mode*, cf. Reißig 2015 for German). This brings us to the last functional category, that of structuring at higher textual levels.

Higher-order functions. In this context, it is important to first mention that there is a certain division of tasks between punctuation units and blank space. In general, it can be claimed for all writing systems that different kinds of blank (or rather unoccupied) spaces organize the writing surface and indicate different levels or classes of textual material (cf. Meletis 2020: Chapter 1.2). However, this structuring can also be achieved—in place of blank spaces or in addition to it—by punctuation, which points to the importance of these higher-order functions.[27] For example, in Amharic, there is (officially) no blank space between words; instead the

[25] This polyfunctionality of a single form is indeed one of the challenges of Bredel's (2008) form-function correlations (cf. Schmidt 2016; see also Nunberg, Briscoe & Huddleston 2002: 1726).

[26] The *gerashyim* ‹״›, which looks like a double quotation mark, marks acronyms; it is placed between the last two graphemes of the non-inflected form of the acronym.

[27] Crucially, spatial arrangement *is* sometimes counted as punctuation in a broader sense, cf. Nunberg, Briscoe & Huddleston (2002: 1725): "[…] the division of a text into paragraphs (marked by a new line and, usually, indentation space) can also be regarded as a matter of punctuation. It is not usual, however, and nor would it be helpful, to extend the domain of punctuation to cover the lay-out of larger units […]."

formally colon-like ‹ः› serves as a word separator.[28] As for larger structures, Thai has units to indicate the end of a long stretch of text ‹๏› or of a chapter or document ‹๚›; similar units also exist in the related Cambodian writing system, which also has ‹᪠› to mark the beginning of literary or religious texts. Other examples include ‹᯼› and ‹᯽› indicating the beginning of different kinds of texts (and genres) in Balinese, ‹꤯› used to mark major sections of text in Manding, and the (obsolete) ‹჻› to mark the end of a paragraph in Georgian.

Interestingly, such units can also be found in the small inventories of Buginese (4 units) and Javanese (5 units), which both also have repetition units. Buginese is especially striking here as it only has a single unit ‹᨞› functionally equivalent to two units of the 'Western' set (full stop and comma). It lacks units for all functions fulfilled by the other units of this set, which implies that the functions of higher-order structuring and repetition may not be 'additional' but rather central in the canon of punctuation inventories worldwide.

4.2.3 Pragmatic functions

Thus far, our focus has been on perception, specifically reading processes. These are clearly primary as we are far more often readers than writers. And while punctuation facilitates reading, it arguably renders writing processes more challenging (Fuhrhop & Peters 2023: 287). However, to be perceived, punctuation must first be produced, which is why we at least want to mention an important pragmatic observation regarding its production here.

Most contemporary writing systems are in some way standardized, making them orthographies (for the difference between the two concepts, see Meletis & Dürscheid 2022: 117). There commonly exist codified norms of varying scope and explicitness, which writers are expected to internalize into their idioscriptual 'system' of writing (cf. Kohrt 1990) and adhere to more or less rigidly depending on the context and register. However, rules may also be underspecified or allow for a choice, resulting in the possibility of licensed variation.[29] And given that orthographic norms are socially but not legally binding, unlicensed variation—both conscious and unconscious—is also possible (cf. Sebba 2007). This is central to what has been called 'digital punctuation', referring to mostly non-standardized uses of punctuation especially in informal digital registers (Androutsopoulos 2020). There, punctuation is (re)purposed pragmatically as it "emerges as a device for organizing written interactions sequentially and establishing shared meanings between participants"; Busch (2021: 2) calls this the 'interactional principle'.

An example is the use of the exclamation mark as part of a so-called *Aufregezeichen* (roughly translated as 'agitation sign') in German (cf. Androutsopoulos 2022); it combines exclamation marks with the digit ‹1› (e. g., ‹!!11!›).[30] Such pragmatic repurposing appears to be widespread, as it has been described for different writing systems: for instance, in Chinese

[28] Amha (2009: 186) notes that in print, the word divider is commonly omitted, meaning word boundaries are indicated by blank spaces (see also Meyer 2016, Ullendorff 1951); by contrast, word dividers are still widely used in handwriting.

[29] An example is the choice of two commas, two dashes or opening and closing parentheses to mark a dislocation in, for instance, German. In other cases, it may also be optional whether or not to punctuate at all, e. g., in the case of the hyphen in long compounds in German.

[30] Functionally, this unit conveys performative anger, with the (supposedly involuntary) production of ‹1› inbetween exclamation marks insinuating that while typing, the writer could not even hold down the shift key on the keyboard.

(as well as Japanese and Korean), the tilde ‹~› is used to convey a friendly tone (cf. Xu & Xia 2023).

The fact that they are so prone to repurposing implies that pragmatic or more generally communicative functions seem to be among the most consciously accessible for language users, in any case more so than units operating at a cognitive level. An example for a unit with such a function is the Amharic sarcasm mark (or *temherte slaq*) ‹¡›, which is supposedly important for genres such as children's literature, satirical comics, and poetry (Tsigie et al. 1999: 6). Indeed, sarcasm and irony have been the target of many invented but largely unofficial punctuation units including the interrobang ‹⁉›. In this context, another interesting instance is the (playful) invention of new punctuation units in the book *Typojis* (Bohatsch 2017). There, all the proposed units have pragmatic functions and, notably, all of them are also 'large' in that they extend vertically over the entirety of the line (cf. Fig. 2), which they have in common with parentheses or the question and exclamation marks. It is also worth mentioning that the label for these units borrows its ending *-oji* from *emoji*, highlighting that we also need to consider the relation between punctuation and emojis as a new type of resource in multimodal writing.

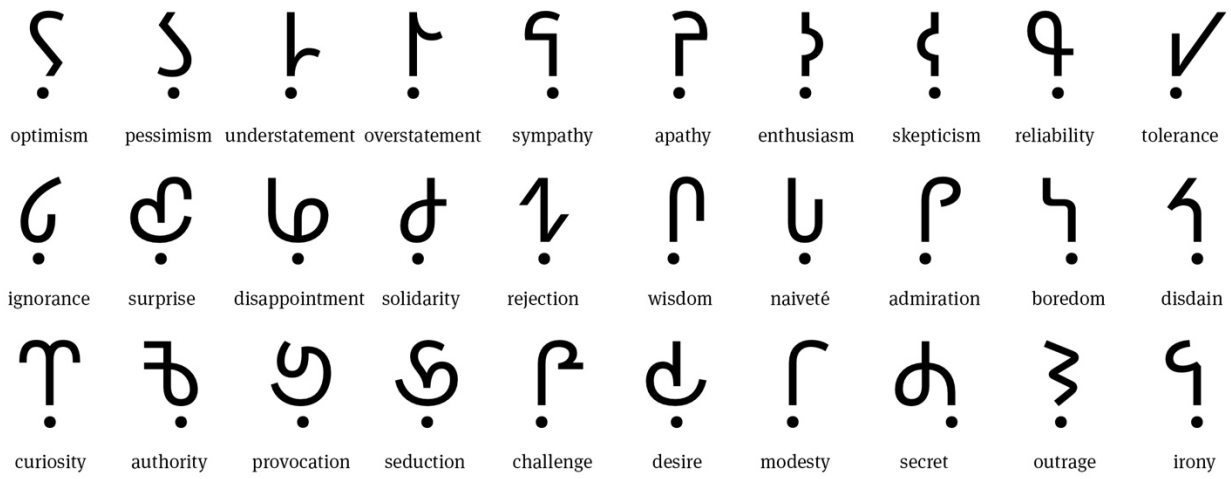

Fig. 2: Typojis with their functions (adapted from typojis.com)

4.3 Forms

Punctuation marks—in this case we can indeed speak of *marks* since the focus is their material appearance—are often relatively small and thus lack graphic[31] salience when compared to their graphic surrounding(s). Several of the 12 punctuation units that serve as the basis for our comparison are small, especially the full stop and the comma, but also the hyphen and the apostrophe.

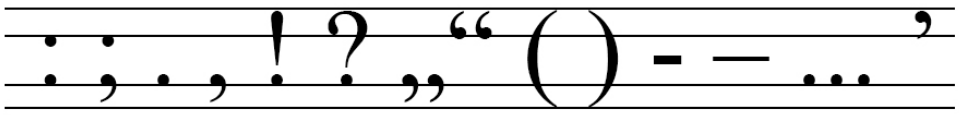

Fig. 3: 'Western' set of punctuation units in the three-space schema

[31] We are not using the term 'visual' here as the lack of salience also applies to braille, a system that is read haptically.

As for their extension as well as position, half of them are located in the lower or middle part of the line,[32] i. e., are not placed in the upper part (unlike ‹'›) and do not vertically fill the line (as done by ‹( )›) (see Fig. 3). This distinguishes them from many default graphemes in the world's writing systems. As for their internal composition, it is noteworthy that half of them consist of multiple discontinuous elements, e. g., ‹;› or ‹"›. While by no means a statistic expression, it can be observed that such a discontinuity—as evidenced by Greek ‹Ξ›, for example—appears to be rare for those shapes of the world's scripts that are employed for default graphemes. Furthermore, associated with the moderate graphic size of punctuation units is the (little) space they require on the writing surface not only vertically but also horizontally:[33] most of them are not full-width,[34] i. e., do not occupy their own segmental space but cliticize onto the graphic units that precede them. These mentioned material and spatial features can be regarded as classificational since they distinguish punctuation units not only from default graphemes but also from digits[35] and special characters.

Notably, this statement is biased alphabetocentrically given that 'smaller' graphemes—sometimes controversially termed 'diacritics'—can also be found in non-alphabetic types of segmental writing systems such as the optional (short) vowel graphemes in the Arabic abjad or the commonly smaller vowel graphemes in abugidas such as Devanagari, which attach to or orbit consonant graphemes. Arguably, on account of their size, their placement relative to as well as their dependence on other graphemes, these vowel graphemes do not represent prototypical default graphemes. Interestingly, this coincides with a functional similarity to punctuation units, as in Arabic, for example, where they are optional, short vowel graphemes are treated as reading aids; however, unlike punctuation units, they have palpable linguistic values as they relate to vowel phonemes.

In the context of lacking graphic salience, Armenian must be mentioned. Not only is ‹՝› small, a unit that is roughly equivalent to a (semi)colon, but also the 'communicative' question ‹՞› and exclamation ‹՛› marks. Graphotactically, they behave like diacritics in that they are placed above (and slightly to the right of) a vowel. Crucially, they are tied to a prosodic function as they modify the tone of the word that they mark. This means that they are not used as sentence-final units but can mark any word in an utterance depending on the writer's intention. The question mark, for example, serves a function analogous to italicization in the following examples: Is *John* going to the cinema? vs. Is John going to the *cinema*?[36]

Given the role punctuation units play in (perceptual) processing, their lack of graphic salience may appear counterproductive. However, for the spoken modality it could be shown

---

[32] In Bredel's featural analysis, they are [¬VERT], i. e., lack verticality.

[33] The interpretation of these terms is relational as it depends on the directionality in a given writing system, i. e., whether lines run horizontally or vertically. The only writing systems in our sample that are nowadays mostly written horizontally but *can* be written vertically are Mandarin and Japanese.

[34] There are exceptions to this, and in Chinese, for example, punctuation units occupy as much space as do default graphemes.

[35] Notably, writing systems in our sample are more likely to have a native set of digits rather than native punctuation: while 15 writing systems use the Arabic digits that also used in English, 11 systems use native digits (Balinese, Bamum, Bangla, Burmese, Gujarati, Hindi, Javanese, Western Kayah, Cambodian, Manding, Thai) and 4 systems use both Arabic and native digits (Amharic, Arabic, Chinese/Mandarin, Japanese). For the native sets, the same can be observed as for punctuation inventories: they are much more visually coherent with respect to the rest of the script than is the case in writing systems using Arabic digits.

[36] See examples at https://www.westernarmenian.me/resources/the-use-of-the-question-mark-in-the-armenian-languagenbsp (March 21, 2024).

that small forms which are clearly distinguishable from default words prevail in usage, as is evidenced by continuers (e. g., *mmhm*) whose spread across languages is commonly described as convergent cultural evolution (cf. Caldwell 2008). And indeed, many units in our sample are not 'small'.[37] They include the Balinese full stop ‹᭽› and comma ‹᭸›, the full stop ‹𖩮› and the colon ‹𖩮𖩮› in Bamum, and the full stop in Bangla ‹।›, Gujarati ‹।›, Hindi ‹।›, and Cambodian ‹។›. As for these native units, it is also interesting to observe this level of variation with respect to the full stop and the comma, units that operate mostly at a cognitive level and aid the parsing of syntactic and textual structures. Being 'larger' means having more graphic salience, of course—it is likely that a large comma is more easily perceived in the periphery of fixations and can thus guide the reading process (and saccade planning) more than a small comma.

Why is it precisely the full stop and the comma that exhibit so much cross-linguistic formal (and possibly to some degree functional) variation? One possible reason is that native forms for these functions had already existed when other units were borrowed from the 'Western' set. This would also explain the instances in which there is a (free or bound) choice between the 'Western' variant vs. the native variant. It is also an important question whether their functions are truly equivalent, as well as whether they are equal with respect to sociolinguistic and pragmatic aspects.

There are two ways in which forms can differ from the 'Western' forms: they can have only a different orientation, or they can exhibit entirely different forms. As for directionality, except for Hebrew, all writing systems with sinistrograde (right-to-left) directionality mirror the forms of 'Western' punctuation units: the question mark is mirrored horizontally ‹؟› whereas the comma ‹،› and the semicolon ‹؛› are mirrored both horizontally and vertically. As for quotation marks, there is a lot of variation regarding the position and directionality of the marks regardless of the directionality of the whole inventory.[38]

Variation can also be found in the positioning of units within the line. Examples of this are the ellipsis in Chinese and Japanese, which is placed not on the base line but in the middle of the line (and reduplicated in Chinese: ‹……›), and the hyphen in Assyrian Neo-Aramaic ‹-› and Hebrew ‹־›, where it 'sits' on the baseline or is positioned higher than the 'Western' hyphen, respectively.

Furthermore, what is striking in the native punctuation inventories in our sample is the graphic coherence of (some of) their forms with the scripts that are employed by the respective writing systems. An example is the inventory of Bamum: consider ‹𖩀 𖩁 𖩂 𖩃 𖩄 𖩅 𖩆 𖩇 𖩈 𖩉 𖩊 𖩋 𖩌 𖩍 𖩎› and ‹𖩮 𖩯 𖩰 𖩮𖩮 - 𖩱›. The first set is a selection of shapes used for (syllabographic) default graphemes, the second is Bamum's punctuation inventory.[39] An assumption as to how this coherence came about is that both the shapes of default graphemes and those of punctuation units were developed simultaneously (and consciously) unlike the 'Western' set that came into

[37] In any case, statements about the 'size' of punctuation units must be taken with a grain of salt, as the internal spatial organization of the line (into vertical subspaces) is of course not the same for all the writing systems in our sample. More information about this can also be found on the respective script pages on r12a.

[38] In sinistrograde Hebrew, there used to be a vertical distinction between opening and closing marks (as nowadays in German): ‹” „›. However, possibly due to influence from English, this vertical distinction was given up, and now both marks are positioned high: ‹” ”›.

[39] Constructional analogies to the 'Western' set can be found in Bamum: the semicolon ‹𖩯› is a combination of the full stop ‹𖩮› and comma ‹𖩰›, the colon ‹𖩮𖩮› a reduplication of the full stop.

use not only after the default graphemes but whose individual units also developed at different points of time. Visual coherence can also be found in mixed sets: in Cambodian the full stop ‹᙮› visually fits in with the shapes of Khmer script, and in Thai, the ellipsis ‹ฯ› shares similarities with the shapes of default graphemes. Such a coherence is notably absent in the writing systems that employ a given script together with (only) 'Western' punctuation units.

Punctuation inventories can also be coherent and systematic in and of themselves, as in Amharic: ‹፠ ፡ ። ፣ ፤ ፥ ፦ ፧ ፨›. They form a set but distinguish themselves markedly from the system's default graphemes. Unlike in the above-mentioned Bamum (and reminiscent of the situation in writing systems using Roman script), the visual features of punctuation units alone indicate a distinction between different functional classes of the writing system.

A strategy that recurs in different inventories is the reduplication of the shape of a unit to indicate a unit operating at a different (often 'higher') functional level: the shape of the comma is reduplicated to indicate a full stop in Balinese, Burmese, and Javanese. In Vai, a reduplicated full stop functions as an exclamation mark, and in Hindi and Gujarati, the full stop can be doubled to mark text boundaries above the sentence level. Reduplication is not to be confused with pairedness as evidenced by ‹( )›; a looser form of such pairedness is constituted by units placed at the beginning and ending of sections such as ‹꧁› and ‹꧂›, which bracket literary/religious texts in Cambodian.

Finally, we find not only the same function in different forms but also the same forms with different functions, a phenomenon we call 'false punctuation friends'. Greek, for example, uses the semicolon ‹;› as a question mark, with a middle period ‹·› taking over the function of the semicolon as we know it from, for example, English. In Armenian, what looks like a period ‹.› is actually the colon, whereas what looks like a colon ‹:› serves as the period.

5. Conclusion and outlook

What our cursory cross-linguistic survey suggests is that variation in punctuation systems may be, on the one hand, mostly graphic, and on the other hand, located mostly at the levels of the linguistic and the sociocultural fits. For example, for certain highlighting purposes, instead of italic type, Chinese uses an emphasis mark (a dot below the highlighted character[s]) since (traditionally) italicization does not exist in Chinese. On the other hand, the sociocultural fit captures that variation is often the result of sociocultural and political factors. It is through this that certain units—such as the inverted question ‹¿› and exclamation ‹¡› marks in Spanish—can also become indexical cultural markers. Depending on underlying and dominant language ideologies, this can either motivate the development and preservation of native units or, as is the case in several writing systems in our sample, lead to their obsolescence in the face of a Westernization of punctuation. Indeed, given pervasive borrowing, punctuation may be a fitting example of cultural diffusion as writing systems deriving from independent inventions of literacy have borrowed (the idea of) punctuation mostly from the same source. Thus, whereas we generally assume a polygenesis of writing, the situation may be special for punctuation, resulting in a predominance of universal traits over diversity in the functionality of punctuation units. This, in turn, is related to the fact that punctuation units cater mostly to the processing fit, and there are only so many cognitive functions that need to be covered; by contrast, default graphemes cater mostly to the structural 'needs' of a language.

Notably, as the fundamtental building blocks of written utterances, default graphemes are, of course, not only a descriptive linguistic phenomenon, while punctuation is not only a cognitive matter. Both complement each other in a writing system that is always simultaneously a linguistic system that must be processed. Thus, methodologically, an exclusively descriptive, offline—i. e., static, language- or structure-oriented—view on punctuation may not be suitable or sufficient, while an 'online' view should also be (at least additionally) adopted for default graphemes. Indeed, for writing in general, but especially for punctuation, structure and processing are so tightly interlocked that grapholinguistics and comparative punctuation research must be interdisciplinary and consider multiple perspectives (cf. Meletis 2024).

These perspectives could be, on the one hand, typological (Which punctuation forms and which punctuation functions occur in different writing systems? Which are unique?) or historical (When and how did punctuation units become established? Why did some, like the leminiscus ‹÷› in German, not persist while others did? What can be deduced from this for possible new punctuation units like ‹*›, which is subject of controversial debates regarding gender-equitable language in German?). On the other hand, comparative punctuation research is also relevant for usage-based approaches to writing. As the use of punctuation units is characterized by variation, we can analyze which phenomena are formally or functionally 'loaded' and at what level of the language system variation plays out. What does it mean if we see that some punctuation units fall out of use while others take on new functions in digital communication? Lastly, these findings can also help us to understand grapholinguistics in general: Which levels of the language system are related to punctuation and how? Why is punctuation so cross-linguistically consistent and why has it persisted in such diverse writing systems? For all these questions it is necessary to clarify the core idea of punctuation. If answered comparatively (as we attempted here), all sides can benefit, because the answers are profitable for both language-specific studies focusing and writing theory and typology as a whole.

References

Amha, Azeb. 2009. On loans and additions to the Fidäl (Ethiopic) writing system. In Alex de Voogt & Irving L. Finkel (eds.), *The Idea of Writing: Play and Complexity*, 179–196. Leiden: Brill. https://doi.org/10.1163/ej.9789004174467.i-396.52

Androutsopoulos, Jannis. 2020. Digitalisierung und soziolinguistischer Wandel: Der Fall der digitalen Interpunktion. In Konstanze Marx, Henning Lobin & Axel Schmidt (eds.), *Deutsch in Sozialen Medien. Interaktiv – multimodal – vielfältig*, 75–94. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110679885-005

Androutsopoulos, Jannis. 2022. Interpunktion und Stilisierung im digitalen Diskurs. Struktur, Registrierung und Pragmatik des ‚Aufregezeichens'. In Florian Busch, Pepe Droste & Elisa Wessels (eds.), *Sprachreflexive Praktiken: Empirische Perspektiven auf Metakommunikation*, 23–49. Berlin: Springer. https://doi.org/10.1007/978-3-662-64597-0_2

Awad, Dana. 2015. The evolution of Arabic writing due to European influence: The case of punctuation. *Journal of Arabic and Islamic Studies* 15, 117–136. https://doi.org/10.5617/jais.4650

Bartsch, Elmar. 1998. Interpunktion. In Gert Ueding (ed.), *Historisches Wörterbuch der Rhetorik*, 515–526. Tübingen: Niemeyer.

Baudusch, Renate. 1980. Zu den sprachwissenschaftlichen Grundlagen der Zeichensetzung. In Dieter Nerius & Jürgen Scharnhorst (eds.), *Theoretische Probleme der deutschen Orthographie*, 193–230. Berlin: Akademie-Verlag.

Baudusch, Renate. 1981. Die Prinzipien der deutschen Interpunktion. *Zeitschrift für Germanistik* 2(2), 206–218.

Berg, Kristian. 2019. *Die Graphematik der Morpheme im Deutschen und Englischen*. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110604856

Bieling, Alexander. 1880. *Das Princip der deutschen Interpunktion nebst einer übersichtlichen Darstellung ihrer Geschichte*. Berlin: Weidmannsche Buchhandlung.

Bodde, Derk. 1991. Punctuation: Its use in China and elsewhere. *Rocznik Orientalistyczny* 47(2), 15–23.

Bohatsch, Walter. 2017. *Typojis: Einige neue Zeichen/A Few More Glyphs*. Mainz: Hermann Schmidt.

Bredel, Ursula. 2008. *Die Interpunktion des Deutschen. Ein kompositionelles System zur Online-Steuerung des Lesens*. Tübingen: Niemeyer. https://doi.org/10.1515/9783484970502

Bredel, Ursula. 2020. *Interpunktion.* 2nd edn. Heidelberg: Winter.

Buchmann, Franziska. 2015. *Die Wortzeichen im Deutschen*. Heidelberg: Winter.

Busch, Florian. 2021. The interactional principle in digital punctuation. *Discourse, Context & Media* 40, 100481. https://doi.org/10.1016/j.dcm.2021.100481

Caldwell, Christine A. 2008. Convergent cultural evolution may explain linguistic universals. *Behavioral and Brain Sciences* 31(5), 515–516. https://doi.org/10.1017/S0140525X08005050

Catach, Nina. 1980. La ponctuation. *Langue française* 45, 16–27. https://doi.org/10.3406/lfr.1980.5260

Evertz, Martin. 2018. *Visual prosody: The graphematic foot in English and German*. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110583441

Fuhrhop, Nanna & Franziska Buchmann. 2009. Die Längenhierarchie: Zum Bau der graphematischen Silbe. *Linguistische Berichte* 218, 127–155. https://doi.org/10.46771/2366077500218_1

Fuhrhop, Nanna & Jörg Peters. 2023. *Einführung in die Phonologie und Graphematik*. 2nd edn. Stuttgart: Metzler. https://doi.org/10.1007/978-3-476-05940-6

Fuhrhop, Nanna, Niklas Reinken & Jonas Romstadt. 2023. Der ‚modalisierende' Gebrauch von Anführungszeichen in Abiturklausuren. *Linguistische Berichte* 276, 405–440. https://doi.org/10.46771/9783967692853_1

Galambos, Imre. 2014. Punctuation marks in medieval Chinese manuscripts. In Jörg Quenzer, Dmitry Bondarev & Jan-Ulrich Sobisch (eds.), *Manuscript Cultures: Mapping the Field*, 341–357. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110225631.341

Gallmann, Peter. 1985. *Graphische Elemente der geschriebenen Sprache. Grundlagen für eine Reform der Orthographie*. Tübingen: Niemeyer. https://doi.org/10.1515/9783111630380

Gallmann, Peter. 1996. Interpunktion (Syngrapheme). In Hartmut Günther & Otto Ludwig (eds.), *Schrift und Schriftlichkeit: Ein interdisziplinäres Handbuch internationaler Forschung*. Vol. 2, 1456–1467. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110147445.2.9.1456

Günther, Hartmut. 1988. *Schriftliche Sprache: Strukturen geschriebener Wörter und ihre Verarbeitung beim Lesen.* Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110935851

Haas, Mary R. 1980. *Thai System of Writing*. 2nd edition. Washington D.C.: Spoken Language Services.

Hamm, John Christopher. 2020. Modern punctuation and layout. In Bruce Rusk, Anatoly Detwyler, Christopher Nugent, Xiao Liu & Jack W. Chen (eds.), *Literary Information in China: A History*, 135–141. New York: Columbia University Press.

Henderson, Leslie. 1985. On the use of the term 'grapheme'. *Language and Cognitive Processes* 1(2), 135–148. https://doi.org/10.1080/01690968508402075

Houston, Keith. 2013. *Shady Characters: The Secret Life of Punctuation, Symbols & Other Typographical Marks*. New York/London: W. W. Norton & Company.
Houston, Keith. 2014. 5 Punctuation Marks That Look Nothing Like They Used To. *HuffPost*. https://www.huffpost.com/entry/post_b_5989482 (18 April 2024).
Jones, Bernard. 1995. Exploring the variety and use of punctuation. In Johanna D. Moore & Jill Fain Lehman (eds.), *Proceedings of the 17th annual conference of the Cognitive Science Society*, 619–624. Mahwah/Howe: Lawrence Erlbaum.
Joyce, Terry & Susanne R. Borgwaldt. 2011. Typology of writing systems: Special issue introduction. *Written Language & Literacy* 14(1), 1–11. https://doi.org/10.1075/wll.14.1.01joy
Klockow, Reinhard. 1980. *Linguistik der Gänsefüßchen. Untersuchungen zum Gebrauch der Anführungszeichen im gegenwärtigen Deutsch*. Frankfurt am Main: Haag und Herchen.
Kohrt, Manfred. 1990. Die 'doppelte Kodifikation' der deutschen Orthographie. In Christian Stetter (ed.), *Zu einer Theorie der Orthographie: Interdisziplinäre Aspekte gegenwärtiger Schrift- und Orthographieforschung*, 104–144. Tübingen: Niemeyer. https://doi.org/10.1515/9783111372280.104
Krahn, Albert E. 2014. *A New Paradigm for Punctuation*. University of Wisconsin-Milwaukee Dissertation. https://dc.uwm.edu/etd/465 (18 April 2024).
Lee, Jeon Kyung. 2014. Korean punctuation systems. *Acta Linguistica Asiatica* 4(1), 29–41. https://doi.org/10.4312/ala.4.1.29-41
Lindbüchl, Isabell. 2014. Ein Komma für den Leser – Sprachverarbeitung und Interpunktion im Deutschen, Englischen und Französischen am Beispiel des Kommas. Diskussionsforum Linguistik in Bayern 4, 69–84.
Meletis, Dimitrios. 2019. The grapheme as a universal basic unit of writing. *Writing Systems Research* 11(1), 26–49. https://doi.org/10.1080/17586801.2019.1697412
Meletis, Dimitrios. 2020. *The Nature of Writing: A Theory of Grapholinguistics*. Brest: Fluxus Editions. https://doi.org/10.36824/2020-meletis
Meletis, Dimitrios. 2021. On being a grapholinguist. In Yannis Haralambous (ed.), *Grapholinguistics in the 21st Century 2020. Proceedings, Part I*, 125–141. Brest: Fluxus Editions. https://doi.org/10.36824/2020-graf-mele
Meletis, Dimitrios. 2024. Schriftlinguistik interdisziplinär, multiperspektivisch, komparativ: Die Erarbeitung struktureller, psycholinguistischer und soziolinguistischer Typologien. In Sabine Krome, Mechthild Habermann, Henning Lobin & Angelika Wöllstein (eds.), *Orthographie in Wissenschaft und Gesellschaft: Schriftsystem – Norm – Schreibgebrauch*, 399–410. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783111389219-022
Meletis, Dimitrios & Christa Dürscheid. 2022. *Writing Systems and Their Use. An Overview of Grapholinguistics.* Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110757835
Meyer, Ronny. 2016. The Ethiopic script: Linguistic features and socio-cultural connotations. *Oslo Studies in Language* 8(1), 137–172. https://doi.org/10.5617/osla.4422
Meynet, Roland. 1971. *L'écriture arabe en question: les projets de l'Académie de langue arabe du Caire de 1938 à 1968*. Beirut: Dār al-Mašriq.
Mora-Marín, David F. 2024. The duplication diacritic: A case study of variation and change in Mayan writing. *Ancient Mesoamerica.* https://doi.org/10.1017/S0956536123000317
Mullaney, Thomas S. 2017. Quote unquote language reform: New-style punctuation and the horizontalization of Chinese. *Modern Chinese Literature and Culture* 29(2), 206–250.
Niikuni, Keiyu & Toshiaki Muramoto. 2014. Effects of punctuation on the processing of temporarily ambiguous sentences in Japanese. *Japanese Psychological Research* 56(3), 275–287. https://doi.org/10.1111/jpr.12052
Nunberg, Geoffrey. 1990. *The Linguistics of Punctuation*. Stanford: CSLI Publications.

Nunberg, Geoffrey, Ted Briscoe & Rodney Huddleston. 2002. Punctuation. In Rodney Huddleston & Geoffrey Pullum (eds.), *The Cambridge grammar of the English language*, 1723–1764. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316423530.021

Parkes, Malcolm B. 1993. *Pause and Effect: An Introduction to the History of Punctuation in the West*. Berkeley/Los Angeles: University of California Press.

Primus, Beatrice. 2007. The typological and historical variation of punctuation systems: Comma constraints. *Written Language & Literacy* 10(2), 103–128. https://doi.org/10.1075/wll.10.2.07pri

Reißig, Tilo. 2015. *Typographie und Grammatik: Untersuchung zum Verhältnis von Syntax und Raum.* Tübingen: Stauffenburg.

Richter, Matthias L. 2023. Towards a broad concept of punctuation. *Bamboo and Silk* 6(2), 159–185. https://doi.org/10.1163/24689246-20230032

Romstadt, Jonas. Accepted. Auslassungspunkte im Korpus. In Nanna Fuhrhop & Laura Scholübbers (eds.), *Interpunktion im Korpus*.

Sampson, Geoffrey. 2016. Typology and the study of writing systems. *Linguistic Typology* 20(3), 561–567. https://doi.org/10.1515/lingty-2016-0027

Schlechtweg, Marcel. 2022. Quotation marks in advanced language acquisition: A reading time experiment on English pure quotation. *Linguistik Online* 118, 99–114. https://doi.org/10.13092/lo.118.9047

Schlechtweg, Marcel & Holden Härtl. 2023. Quotation marks and the processing of irony in English: evidence from a reading time study. *Linguistics* 61(2), 355–390. https://doi.org/10.1515/ling-2021-0079

Schmidt, Karsten. 2016. Der graphematische Satz: Vom Schreibsatz zur allgemeinen Satzvorstellung. *Zeitschrift für germanistische Linguistik* 44(2), 216–256. https://doi.org/10.1515/zgl-2016-0011

Schmidt, Karsten. 2018. *Phonographie und Morphographie im Deutschen. Grundzüge einer wortbasierten Graphematik*. Tübingen: Stauffenburg.

Sebba, Mark. 2007. *Spelling and Society: The Culture and Politics of Orthography Around the World*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486739

Stetter, Christian. 1997. *Schrift und Sprache*. Frankfurt am Main: Suhrkamp.

Ström Herold, Jenny & Magnus Levin. 2021. On brackets in translation (or how to elaborate in brackets). *Bergen Language and Linguistic Studies* 11(1), 121–144. https://doi.org/10.15845/bells.v11i1.3441

Stryjewska, Anna. 2016. Punctuation, modern. In Rint Sybesma (ed.), *Encyclopedia of Chinese Language and Linguistics.*

Tiffin-Richards, Simon & Sascha Schroeder. 2018. The development of wrap-up processes in text reading: a study of children's eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44(7), 1051–1063. https://doi.org/10.1037/xlm0000506

Tsigie, Asteraye, Berhanu Beyene, Daniel Aberra & Daniel Yacob. 1999. A roadmap to the extension of the Ethiopic writing system standard under Unicode and ISO-10646. Paper presented at the 15th International Unicode Conference, San Jose, CA, August/September 1999. http://yacob.org/papers/DanielYacob-IUC15.pdf (22 March 2024)

Twine, Nanette. 1984. The adoption of punctuation in Japanese script. *Visible Language* 18(3), 229–237.

Ullendorff, Edward. 1951. Studies in the Ethiopic syllabary. *Africa: Journal of the International African Institute* 21(3), 207–217. https://doi.org/10.2307/1156593

Veith, Werner H. 1985. Graphem, Grapheotagmem und verwandte Begriffe. In Gerhard Augst (ed.), *Graphematik und Orthographie. Neuere Forschungen der Linguistik, Psychologie und Didaktik in der Bundesrepublik Deutschland*, 22–43. Frankfurt et al.: Peter Lang.

Xu, Huan & Dengshan Xia. 2023. Digital tildes (“~”) may convey more: analyzing innovative uses of tildes in Chinese WeChat messages. *Language and Semiotic Studies* 9(3), 443–460. https://doi.org/10.1515/lass-2023-0009

Appendix

Tab. 3: Comparative survey of 30 punctuation inventories

Writing system: Language (Script)[1]	Punctuation												
	.	;	,	:	-	–	…	’	?	!	( )	“”	Notes
Amharic (Ge’ez) ABJ	። .	፤	፣	፥ ፦	-	–	…		?፧	!	( )	«»	‹፨› as a paragraph marker, ‹፠› as a section marker, ‹¡› as a sarcasm mark, all three not commonly used; ‹፧› used as question mark, ‹፡› as a word separator (mostly used in handwriting)
Arabic ABJ ⟵	.	؛	،	:	-	–	…		؟	!	( )	“”	‹؍› date separator
Armenian ALPH	։ .	․	,	՝	- ‿		…	’	՞	՛	( )	«»	‹’ › only used in Western Armenian, ‹՞› and ‹՛› modify vowel (placed above and slightly to the right of vowel whose sound is being modified), emphasis mark ‹՛› behaves the same way
Assyrian Neo-Aramaic (Syriac) ABJ ⟵	.	؛	،	܆ ܇	-				؟		( )		‹܏› as an abbreviation mark (used horizontally above abbreviation), ‹-› sits on the baseline
Balinese[†] ABU	᭟		᭞	᭝									‹᭚› used at the beginning of a letter, story, or verse; ‹᭛› used at the beginning of (mostly religious) texts; ‹᭜› used for line breaking
Bamun (Bamum) SYLL	꛳	꛶	꛵	꛴	-				꛷				‹꛲› used to introduce proper nouns; can be used to disambiguate homographs
Bassa (Bassa Vah) ABU	+ .		,		-				?			“”	‹+› used as an alternative to the Western full stop
Bangla (Bengali) ABU	. ।	;	,	:	-			’	?	!	( )	“”	‹◌৪› marks initial abbreviations, ‹৺› is used alongside the name of deceased persons
Buginese[i] (Lontara) ABU	᨞		᨞										‹᨞› functionally inbetween comma/full stop; ‹᨟› used to indicate the end of a section; ‹ᨗ› indicates doubling of root or word
Burmese[p] ABU	။		၊	:					?		( )	“”	
Cherokee SYLL	.	;	,	:					?	!	( )	“”	
Chinese/Mandarin (Han	。	；	， 、	：	-	—	……		？	！	（ ）	“”	all units are fullwidth; ‹、 › used for enumerations; different forms

[1] If only one term is given, this means that the language’s and the script’s name is identical: i. e., the writing system for Arabic uses Arabic script.

Simplified)† MORPH													of quotation marks in vertical writing
Cree Plains (UCAS[2]) ABU	᙮ .	;	,	:	- =				?	!	( )		
Dhivehi (Thaana) ABU ⟵	.	؛	،	:					؟	!	( )	”“	
English (Roman) ALPH	.	;	,	:	-	– —	…	’	?	!	( )	“”	
Georgian ALPH	.	;	,	:	-				?	!	( )	„“	‹:› formerly used to indicate the end of a paragraph
German (Roman) ALPH	.	;	,	:	-	–	…	’	?	!	( )	„“	
Greek ALPH	.	·	,	:	-	—	…		;	!	( )	«»	‹/› can be used to indicate common abbreviations
Gujarati ABU	. ।	;	,	:				’	?	!	( )	“”	‹ ॥› marks boundaries of text above the sentence level, ‹°› used to indicate abbreviations
Hebrew ABJ ⟵	.	;	,	:	־				?	!	( )	” ”	‹”› and ‹’ › used to mark acronyms and abbreviations (like abbreviation period), quotation marks ‹””› have the same form in both positions, ‹־› is positioned higher than Western hyphen
Hindi (Devanagari) ABU	. ।	;	,	:	-				?	!	( )	“”	‹ ॥› marks boundaries of text above the sentence level, ‹°› used to indicate abbreviations
Japanese (kanji, hiragana, katakana, Roman)† SYLL, MORPH	。		、 ,	:	= -	–	… …		?	!	( )	「 」	all units are fullwidth; ‹·› used for short inline lists; different forms of comma and hyphen used in vertical vs. horizontal writing
Javanese[p] ABU	꧉		꧈	꧇							꧌ ꧍	꧊	‹꧋› used to introduce a paragraph or a section; ‹꧈› is used for acronyms, ‹ꧏ› indicates the repetition of a syllable
Western Kayah (Kayah Li) ALPH	꤯	;	,	:					?	!	( )	“”	
Cambodian (Khmer)[p] ABU	។			៖	-				?	! ◌៎	( )	“”	‹-› used between different parts of names; ‹៕› used at the end of a chapter/a text; ‹៙› marks the beginning of literary/religious texts, ‹៚› marks the end of these texts; ‹◌៎› rarely used over final

[2] Unified Canadian Aboriginal Syllabics.

													consonant of a word for emphasis, ‹ᩅ› as a repetition sign
Korean (Hangul) ALPH	.	;	,	:	-	–	…		?	!	( )	“”	‹·› used to disambiguate or in short inline lists; comma, period, and parentheses have different forms in vertical writing: ‹、 › ‹ 。 › ‹ 『』 ›
Manding (N'Ko) ALPH ⟵	.	؞	، ؛	:	-				؟	⁝	( )	»«	‹𞤮› used to end major sections of the text; ‹،› and ‹؛› can be used distinctively within the same text
Russian (Cyrillic) ALPH	.	;	,	:	-	—	…		?	!	( )	«»	
Thai[p] ABU	.		,	:	-	–	ฯ …		?	!	( )	“”	space used as phrase rather than a word marker; ‹ฯ› marks end of a long segment of text, ‹๛› marks the end of a chapter or document, ‹ๆ› marks repetition of preceding letters, ‹ฯ› as an abbreviation/elision marker
Vai[i] SYLL	. *	;	, ꘍	:					? ꘏	! **	( )		

† no spacing

[i] inconsistent use of spacing between words (or larger units)

[p] spacing between phrases rather than words

ALPH alphabet, ABJ abjad, ABU abugida, SYLL syllabary, MORPH morphographic

⟵ written right to left (if not noted: nowadays written left to right)