



The grapheme as a universal basic unit of writing

Dimitrios Meletis

To cite this article: Dimitrios Meletis (2019) The grapheme as a universal basic unit of writing, Writing Systems Research, 11:1, 26-49, DOI: [10.1080/17586801.2019.1697412](https://doi.org/10.1080/17586801.2019.1697412)

To link to this article: <https://doi.org/10.1080/17586801.2019.1697412>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 09 Dec 2019.



Submit your article to this journal [↗](#)



Article views: 4242



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)

The grapheme as a universal basic unit of writing

Dimitrios Meletis 

Department of Linguistics, University of Graz, Graz, Austria

ABSTRACT

The grapheme appears to be a central concept of grapholinguistics. However, there is no consensus on how it should be defined. Some use the concept of grapheme in their work but fail to give a definition while others altogether reject it. When the concept is defined, it is interpreted either as a written unit which refers to a phoneme (this is termed the *referential view*), or as a written unit that is lexically distinctive (*analogical view*), which is tested via written minimal pairs such as <house> and <mouse> analogously to phonological minimal pairs which can be used to discover phonemes. A problem of these two views is that they are restricted to alphabets. A universal conception of the grapheme inclusive of all types of writing systems would make possible the uniform description and, consequently, the comparison of diverse writing systems. Such a conception is proposed here: Graphemes are units of writing which are (1) lexically distinctive, (2) have linguistic value (mostly by referring to phonemes, syllables, morphemes, etc.), and are (3) minimal. These criteria are characterised in detail, and examples from writing systems such as Arabic, Chinese, Devanāgarī, German, Japanese, Korean, Tamil, and Thai highlight their cross-linguistic applicability.

ARTICLE HISTORY

Received 29 April 2019

Accepted 19 November 2019

KEYWORDS

Grapheme; grapholinguistics; comparative graphematics; grapholinguistic terminology; comparative concept

1. Introduction

Every writing system is structurally complex. In every writing system, regardless of how much it differs from other writing systems, there exists some kind of minimal unit with a linguistic function that is used compositionally to build larger units. Irrespective of which exact functions it fulfils, this minimal unit represents the basic entity of writing systems. In analogy to basic linguistic units such as the phoneme or morpheme, it has been proposed that it be called *grapheme*. So far, so good. A number of scholars across disciplines have used this term, but often have done so vaguely, failing to make explicit what they mean by it.¹ The instances in which *grapheme* is indeed defined reveal that understandings of the concept diverge considerably, reflecting a lack of consensus over what the term signifies.

More than three decades ago, Kohrt (1986, p. 81), author of an extensive treatment of the grapheme (Kohrt 1985), noted: '[P]erhaps you are convinced that the problem of providing [the grapheme] with an adequate definition has already been solved or is simply insignificant', concluding: 'As for me, I share neither of these views'. To this day, the matter remains relevant but unsettled. Despite this, there currently appear to be no aspirations to negotiate a definition of *grapheme* that researchers concerned with writing would agree on. Quite to the contrary: the terminological and

CONTACT Dimitrios Meletis  dimitrios.meletis@outlook.com  Department of Linguistics, University of Graz, Merangasse 70/III, A-8010 Graz, Austria

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

conceptual chaos surrounding the grapheme has led some eminent scholars of writing to reject the concept altogether, resulting in a depreciation and, consequently, a palpable stagnation of any fruitful discussion surrounding it.

At the same time, however, the popularity of research on writing systems is rising, and grapho-linguistics is gaining currency as a linguistic subdiscipline, having gradually been emancipated from its initial focus on alphabetic writing systems and expanded in its scope to include all types of writing. As much as it appears obvious that all types of writing systems exhibit basic units, it is unclear what the common core of a generalised basic unit could be. ‘What is the grapheme?’ is thus no simple question, which makes necessary a re-evaluation and definition of the concept.

Instead of *grapheme*, what is currently predominant is the use of more or less established terms such as *letter*, *character*, or vague placeholders like *sign (of writing)*, and *symbol*. These terms all lead double lives as lay terms and quasi-technical terms, and their use is problematic. For example, labelling the smallest functional unit of the German writing system ‘a letter of German’ might be terminologically adequate in an isolated description of the German writing system. It might even allow for comparisons with other writing systems of the same type, i.e. alphabets, writing systems in which both consonant and vowel phonemes are graphematically represented by units of equivalent status, and comparisons of these units, i.e. the alphabets’ respective ‘letters’. Ultimately, however, this conceptual and terminological choice dissociates the results of such an alphabet-specific description from a global theory of writing by obscuring the fact that ‘a letter of German’ and ‘a character of Chinese’ might share features which allow for a unified classification of these units as graphemes of their respective writing systems. In general, comparative analyses across various types of writing systems become more difficult without a shared conceptual and terminological framework, which might be the main reason why they are so rare.² Comparison, however, is the crux of an emerging theory of writing (Meletis, 2018), and in order to compare diverse writing systems, we are in need of a benchmark. Therefore, positing a concept of *grapheme* and defining it in a broad manner, thereby making it applicable to all writing systems and facilitating comparisons between them, represents an important and promising endeavour for the advancement of grapho-linguistics.

In this paper, first, the most important views on the grapheme will be presented, and on this basis, a new and improved definition of the concept will be proposed. Crucially, unlike past alphabetocentric definitions, the suggested conception of the grapheme takes into consideration all types of writing systems. Examples from a variety of systems will demonstrate this.

2. Strategies of dealing with the grapheme

The term *grapheme* made its first appearance at the beginning of the twentieth century. Jan Baudouin de Courtenay is commonly credited with coining it³ and interpreting the associated concept as the psychological representation of ‘letters’ (see Kohrt, 1985, pp. 171–172 for a list of differing but conceptually related uses of the term in Baudouin de Courtenay’s works). Crucially, early definitions of the concept were almost always linked to works of phonology, and ‘[e]very conception of the “grapheme” had some previous interpretation of the term “phoneme” which served as a model’ (Kohrt, 1986, p. 82; see also Section 2.4.). The concept’s further development moved into two separate directions closely linked to the respective grapho-linguistic tradition in question: in Anglo-American research on writing systems, the term was largely abandoned, and it is absent in milestone works such as Gelb’s (1969) *A study of writing* and plays a negligible role in modern works such as Rogers (2005) and Sampson (2015). By contrast, in the German-speaking tradition, the grapheme quickly advanced to a central subject of discussion when German grapho-linguistics flourished in the 1980s and was eventually joined by proposals of other graphematic units such as the graphematic syllable (Fuhrhop & Buchmann, 2009), the graphematic foot (Evertz, 2018), the graphematic word (Fuhrhop, 2008), and the graphematic sentence (Schmidt, 2016).

Overall, four major strategies of dealing with the grapheme have emerged: rejecting the grapheme, not defining the grapheme, preferring the concept of *letter* over *grapheme*, and, most importantly, assuming and defining a concept of *grapheme*. While the first two strategies are characteristic of the Anglo-American literature, the latter two are closely associated with German grapholinguistics. The following subsections consider each of them in turn.

2.1. Rejecting the grapheme

The first and most radical of the strategies is the rejection of the grapheme. The most prominent figure opposing the idea of a grapheme is Peter T. Daniels. For him, the grapheme ‘has become nothing more than a pre-theoretic, fancy, scientific-sounding word for “letter” or “character” and ought not to be part of technical discourse’ (Daniels, 2017, p. 88). One of the main reasons for this assessment that he lists in a number of publications is the different status of language and writing. In this context, Daniels (2017, p. 88) notes that phonemes and morphemes are unconscious properties of language, whereas writing is ‘not an unconscious, built-in feature of a mind’ and therefore ‘cannot a priori be assumed to be analyzeable in a parallel way’ to language (Daniels, 2017, p. 88). This constitutes Daniels’s major contention (dating back to Daniels, 1991; cf. Herrick, 1994 for a rebuttal) not only against a grapheme, but against *graphem(at)ics* as a subbranch of linguistics, i.e. a branch that treats writing systems as linguistic systems, including units of writing which are assumed in analogy with linguistic units such as the phoneme or the morpheme. Note, however, that graphematics is a well-established and accepted field in the German grapholinguistic tradition and has been for a couple of decades, at least since the 1980s (Berg & Evertz, 2018; Dürscheid, 2016, Chapter 4; Fuhrhop & Peters, 2013). This aptly underlines the unfruitful divide between different grapholinguistic traditions and the lack of mutual reception of their respective research.

Many statements Daniels makes about writing and its differences from language are accurate: writing is not acquired without instruction and is, in fact, both phylogenetically and ontogenetically secondary to speech, and not only speech, but also sign language (see also Dürscheid, 2016, p. 30). There exists no brain region that has evolved specifically for reading or writing, as areas with related functions have rather been neuronally recycled to take on these tasks (Dehaene, 2009, 144–147). Finally, as mentioned above, writing is more conscious than language as writers and readers need to be – at least in the process of acquiring literacy – consciously aware of the units of writing they employ (Kohrt, 1986, p. 93). Indeed, readers and writers are not only conscious of the units of writing, but, by virtue of using them, must also become aware of the linguistic units that these units of writing refer to. Writing requires not only phonological and morphological awareness, but, more generally, metalinguistic awareness. Writing, which, unlike speech, is an optional modality of language,⁴ always constitutes an analysis of the underlying language system. Writing is not only conscious itself – it also makes language conscious.

Should the fact that writing operates at another level than language, making the unconscious properties of language conscious, preclude us from analysing it with the same or similar tools and concepts that are used to analyse subsystems of language such as phonology and morphology? Rogers (2005, p. 11) denies this, claiming ‘the fact that the data of language and writing are different in nature does not preclude our using a similar theoretical framework’ (see also Primus, 2004, p. 237). A reservation about Daniels’s rejection of graphematics that cannot easily be resolved is: is writing actually always so conscious? Once acquired successfully, are the processes of reading and writing not often dominated by automatisms?

In any case, what must be carefully separated is the discussion of the grapheme as a unit relevant in processing on the one hand, which is addressed by this question of consciousness, and the assumption of the grapheme merely as a unit of description on the other which remains unaffected by this question. The present paper is focused on the latter, the grapheme as a unit

of description. Whether such a descriptive concept is psychologically real must be evaluated in a separate step.

With respect to the distinction between an etic and an emic level, it is, contrary to Daniels's claim, undeniable that there is an emic level in writing,⁵ as concrete, graphetic substantiations can be – even must be – classified into abstract emic units to make an analysis of writing possible. The emic units relevant to writing are, at a graphetic level, *basic shapes*,⁶ and, at a graphematic level, *graphemes*. This is in accordance with the definition of emic units as 'invariant form[s] obtained from the reduction of a class of variant forms to a limited number of abstract units' (Nöth, 1990, p. 183).

One of Daniels's more trivial objections is of terminological nature: together with his collaborator David Share, he claims that "[g]rapheme" has had so many different interpretations that in writing systems theory it is meaningless' (Share & Daniels, 2016, p. 23). While the first part of this statement is true, it is not causally linked to the second part: different interpretations of the term are not a legitimate reason to completely discard the term or, more importantly, the idea behind it. One of the many conceptions of the grapheme that circulate could be right, or the right one has yet to be formulated. The truth, I believe, lies somewhere in the middle: some of what has been proposed under the heading of *grapheme* will turn out to be useful while other aspects must indeed be dismissed.

What Daniels's issue with the grapheme reflects is that many efforts in defining it have taken place in isolation from the larger linguistic picture. Phoneme and morpheme definitions, let alone the definitions of a large number of other concepts and categories needed in the description of language systems or linguistic data necessarily ventured for universality fairly quickly. By contrast, for the concept of grapheme, it was always 'the *grapheme* in German' or (more seldom) 'the *grapheme* in Chinese', but never a more generally defined concept of *grapheme* that can explain how *all* writing functions at its core. There is no textbook definition of *grapheme*. The specificity of different circulating definitions of *grapheme* is, at least to some degree, an understandable reaction to the typological diversity of writing systems which, at first glance, might discourage grapho-linguists from seeking out the possibly universal nature of the grapheme. Even Kohrt (1986, pp. 91–92), who has thoroughly studied the history of the concept of *grapheme* (Kohrt, 1985), does not believe 'that "logograms", "morphograms" and "phonograms" are to be considered as different appearances of one and the same kind and that they should be subsumed under the notion "grapheme"'. However, the visual diversity of the world's scripts⁷ and the functional diversity of the world's writing systems are deceptive: at the very core, there are sufficient similarities across these systems that allow assuming a universal grapheme concept.

2.2. Not defining the grapheme

Numerous authors of grapho-linguistic works who use the term fail to define what they mean by *grapheme*. Essentially, they leave it up to the reader to decipher its meaning in a given context. This empty and careless use is indeed so frequent that Daniels's above-mentioned aversion to the term becomes understandable (although, admittedly, none of this is the term's fault). This type of vague use is predominantly found in non-linguistic works, e.g. works of psychology, the cognitive sciences, pedagogy, speech technology, etc., but also in works from linguistic subfields that are not primarily focused on writing. Alas, given the interdisciplinarity of grapho-linguistics, as they address writing, these works become relevant to the field, which is how their casual use of *grapheme* becomes a problem.

Take, for example, a study about handwritten character production in Chinese. In the first paragraph of the paper, the authors state that '[t]he letters or graphemes serve as functional units in the orthography of a word' (Chen & Cherng, 2013, p. 1). Here, evidently, *grapheme* is used as a synonym for *letter*, supporting Daniels's statement that *grapheme* often serves as a more scientific-sounding alternative. In a study about reading in Thai, Winskel and Iemwanthong (2010,

p. 1024, p. 1028) write of, for example, ‘consistent grapheme to phoneme mapping’ or ‘children’s phonological knowledge and ability to map sublexical units onto graphemes’. They fail to define what a grapheme is or what units comprise the grapheme inventory of Thai. *Grapheme* is obviously meant to be some sort of functional written unit, but in a typologically phonographic but non-alphabetic writing system (an *abugida*, Daniels, 2017) such as Thai, is it so intuitively obvious what a grapheme is? Is it just the consonant ‘signs’ or also the secondary and dependent vowel ‘signs’ (cf. Section 3.3.)? Taha (2013) writes that ‘Arabic is a language written in an alphabetic system of 29 letters [...]. Arabic letters have more than one written form, depending on the letter’s place in a word: beginning, middle, or end’ (Taha, 2013, p. 725). Here, *letter* seems to be an abstract notion whose materialisation are *written forms* that depend on the position of the letter in a word. Later, Taha changes his terminology, writing ‘the basic and non-connected shape of the grapheme س could change according to its placement within the word’ (Taha, 2013, p. 725). Letters are now *graphemes*, written forms are now *shapes*.

There are many more examples like these. They reveal that *grapheme* has no fixed meaning, but, more deleteriously, they also evoke the impression that there possibly cannot be a coherent and cross-graphological definition of it.

2.3. Preferring the concept of letter

Some scholars do not dismiss the grapheme altogether, but avoid it in favour of other, supposedly more specific terms. In his seminal work on the graphematics of German, Martin Neef (2005) opts out of using *grapheme* and relies on *letter* (German *Buchstabe*) instead. However, unlike Chen and Cherng (2013), he does not treat these terms as synonyms, and unlike Daniels, he does not argue that there is no grapheme. Instead, he claims that the concept of *letter* is sufficient for the description of the German writing system, a context in which, while it might be a useful concept elsewhere, the grapheme is superfluous. With respect to a prospective comparison of typologically diverse writing systems, a critical evaluation of the choice of *letter* as the smallest functional unit of writing is in order.

Neef (2005, p. 37) uses the German word <Schnee> ‘snow’ to illustrate that *grapheme* is a notion different from *letter*. Since he analyses <sch> as one grapheme, <Schnee> consists of the four graphemes <sch>, <n>, and two instances of <e>, and simultaneously of six letters, as <sch> is itself a complex grapheme made up of the three letters |s|,⁸ |c|, and |h|. Since the smaller constituents of <sch>, i.e. <s> as well as <h> and <ch> (see below) are also commonly analysed as graphemes, Neef notes that in a model that includes graphemes, *grapheme* must be defined in a way that accounts for the possibility of graphemes being comprised of one letter, as in <s>, or a combination of letters, as in <sch>. A theory of graphematics, he concludes, can function with the unit *letter* alone or with the units *letter* and *grapheme*. However, due to graphemes consisting of multiple letters such as <sch>, such a theory can never function with the grapheme alone. For reasons of economy, Neef decides *letter* is sufficient for his description of German graphematics. This, then, leaves open the question of how complex units such as <sch> are to be treated in such a *letter*-based model. Neef calls these units *fixed letter combinations* (German *feste Buchstabenverbindungen*, Neef, 2005, p. 41). Also, since *letter* already functions as the smallest graphematic, i.e. abstract linguistic unit, Neef is in need of a new designation for the material manifestation of this unit, what I call *basic shape*. He opts for *letter body* or *letter gestalt* (translated from German *Buchstabenkörper*, Neef, 2005, p. 39).

Conceptually, I agree with Neef. I merely propose a terminological shift. The reason for my preference of *grapheme* is that a shared term – which must be accompanied by a shared definition and concept – allows comparisons between different writing systems. While *letter* or *character* might suffice in individual descriptions, a definition of *grapheme* that captures the minimal linguistically functional unit of any given writing system reflects that at the core, they all share a crucial function: visually representing language (and this is not restricted to speech). As mentioned above,

Neef’s introduction of *letter body* proves that his functional interpretation of *letter* leaves a vacancy regarding the graphetic units that take part in graphematic relations. By contrast, I argue that *letter* is a graphetic term, as is *character*. Both of these terms signify types of basic shapes and can be replaced by the umbrella term *basic shape*. The additional information these terms give in comparison with the general *basic shape* is the script they stem from, although this is only partially true for *letter*, which enjoys great popularity whenever talk is of the units of an alphabet – not only of alphabets using Roman script, but also the Georgian, Armenian, Cyrillic, and Greek scripts. This means *letter* is indeed used graphematically, as the designation for a grapheme that is specific to a given type of writing system – the alphabet. However, *letter* has spread even wider and become the designation of units that are used in non-alphabetic writing systems: take the Arabic script, mostly employed for *abjads*, whose units are also sometimes referred to as *letters* (Saiegh-Haddad & Henkin-Roitfarb, 2014, p. 15), which, however, is not problematic *per se*, since the units of *abjads* also represent phonemes, even if – in contrast to alphabets – typically only consonant phonemes (and long vowel phonemes) rather than both consonant and vowel phonemes. This overgeneralisation of *letter* reveals aspirations of finding a common ground across writing systems by using a uniform designation for their respective basic units. It is this very motivation that also underlies the attempt of finding a universal definition of *grapheme*.

Another model that does not dispense with the grapheme but reduces its relevance is the suprasegmental model of writing developed by Beatrice Primus and her colleagues (cf. Figure 1). It originally evolved from the idea that the syllable can be described not only as a unit of speech but also as a unit of sign language and written language. As such, *syllable* is interpreted as an amodal concept, with the defining criterion being an alternation between salient and non-salient units⁹ (Primus, 2003).

As in Neef’s model, *letter* is interpreted as a graphematic and central unit in the suprasegmental model (Berg, 2019, Chapter 3.1). Additionally, the suprasegmental model assumes a level of skeletal positions above the level of letters. The authors suggest that these skeletal positions could possibly be graphemes. In the majority of cases, one letter is associated with one skeletal position, i.e. single letters are often simultaneously graphemes. By giving letters priority, Primus and her colleagues imply – in a manner similar to Neef’s – that in many cases, as letters coincide with graphemes, the concept of *letter* actually does suffice. The grapheme is only of interest for letter

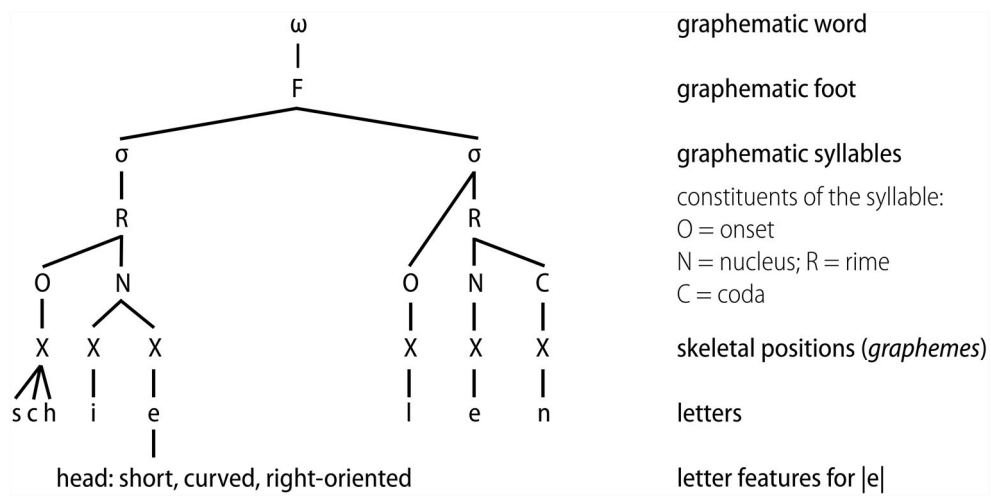


Figure 1. Suprasegmental model of writing, illustrated with the German example <schielen> ‘squint’, adapted from: Berg, Primus, & Wagner (2016, p. 351).

combinations that behave like single letters, for example the above-mentioned <sch> in German (Schmidt, 2018, p. 138). In these cases, letters are combined, and these letter combinations are associated with a single skeletal position (cf. Figure 1), which is how in this model, too, <sch> is regarded as a single grapheme consisting of three letters (Berg, Primus, & Wagner, 2016, p. 351). Unlike Neef's model, in which these sequences are called *fixed letter combinations*, the suprasegmental model initially called them *graphemes*; however, this was abandoned in newer versions of the model.¹⁰

The reliance on the letter as the most central unit in these otherwise promising approaches is – given their scope and interest – understandable, but, at the same time, remains problematic as it deflects from the fact that non-alphabetic systems might share features with alphabetic systems and could be modelled in similar ways (Meletis, 2017, p. 112).

2.4. Proposing a grapheme

In the productive German grapholinguistic tradition, many authors have posited the grapheme as a unit relevant for the description of writing systems. However, even among these scholars who agree on the existence of a grapheme, there is no consensus about what exactly it is. The two most influential conceptions of the grapheme are the *referential view* and the *analogical view* (Kohrt, 1986; Lockwood, 2001).¹¹ They are closely linked to differing views on the relation between writing and speech. While scholars adhering to the referential view interpret writing as a representation of speech (or, more precisely, phonology; this view of writing being dependent on phonology has been called the *dependency hypothesis*), proponents of the latter view see writing as a modality of language relatively autonomous from speech (or, again, phonology; this has been called the *autonomy hypothesis*, Dürscheid, 2016, pp. 35–41).

In the referential view, graphemes are treated as signs of phonemes, whereas in the analogical view, graphemes are identified analogously to phonemes and are defined as the smallest functional units of writing, regardless of whether they correspond with or represent phonemes (or other units of language such as morphemes). Like Kohrt (1986, p. 89), I argue that in isolation, both views prove insufficient for explaining what the grapheme is – at least cross-linguistically. However, I also agree with Lockwood (2001, p. 307) in that ‘students of writing need to include both kinds of relations in any model they adopt’. Accordingly, neither of them gets it completely right, but there is some truth in both of them.

Proponents of the referential view treat graphemes as units that stand for phonemes, as “phoneme signs” (Kohrt, 1986, p. 84). This has been fervently criticised. As Günther (1988, p. 76) points out, if *graphemes* were derived from phonemes, there would not be a need for the concept of *grapheme* to begin with, as graphemes would simply be written labels for phonemes. Graphemes, then, would not be units, but relations, correspondences, or rules. Another issue that Günther addresses is allography. In the referential view, allographs are units that are used to write the same phoneme, which leads to the absurd situation of assigning allographs to a phoneme instead of a grapheme (Günther, 1988, p. 76). For example, allographs that Garbe (1985, pp. 12–13) lists for the phoneme /f/ in German are <f, ff, v, w, fe, ph>. This unveils problems that the referential view is confronted with: Firstly, the direction of analysis is from phoneme to grapheme. Primacy is given to encoding phonology in written form, which, in processing terms, translates to production processes. This is not a problem *per se*, but a choice – a choice, however, that yet again gives primacy to speech and thereby strongly insinuates that writing is merely a derivation of speech when in fact it is much more. Modern approaches either choose the opposite direction for their analysis, i.e. from grapheme to phoneme, or consider both by postulating bidirectional mapping relations (Evertz, 2016, p. 381).

Also, a form such as <ff> is not, as Garbe proposes, an allograph of a phoneme /f/ or even of a grapheme <f>. As graphemes are to be defined as *minimal* functional units of writing (cf. Section 3.1., criterion (3)), it is rather a sequence of two functional units, two instances of the grapheme

<f>. The doubling is determined by higher levels in the hierarchy of graphematic units (Birk, 2013; Schmidt, 2018, pp. 32–33), in this case the syllable level: in <Löffel> ‘spoon’, for example, the doubling of <f>, in the referential view, signifies the phonological quantity of the preceding vowel and the fact that the /f/ that <ff> stands for is ambisyllabic (Primus, 2010, pp. 20–25). In the analogical view, specifically in non-linear suprasegmental models such as the one outlined above, the doubling is determined by the structure of the graphematic foot and signifies that the preceding <ö> is associated with only one (instead of two) skeletal positions (Evertz, 2018). Thus, it is the structure of graphematic feet and graphematic syllables in German that conditions the variant <ff>. In both the referential view and the analogical view, <ff> is not an idiosyncratic spelling, but an explainable one (Fuhrhop & Peters, 2013, pp. 229–238). Thus, there is no need to assume it as a separate unit.

The second influential conception of the grapheme in German grapholinguistics is the analogical view. Adherents of this view treat the grapheme as the smallest distinctive unit of writing (cf. Fuhrhop & Peters, 2013, p. 202; Günther, 1988, p. 77; Rogers, 2005, p. 10). Here, the phoneme does not serve as the unit that the grapheme refers to, but instead as a methodological model of how distinctive units can be discovered. Consequently, the analogical view is not structurally, but methodologically dependent on phonology (Berg, 2019, p. 26). In structuralism, the phonemes in a language’s phoneme inventory are identified with the help of substitution, which potentially yields distinctive minimal pairs, even if nowadays the structuralist view of the phoneme has become outdated. The same, analogists argue, can be done in written language. Minimal pairs such as <house> and <mouse> reveal that the contrasting units, if they render a difference in meaning and create two existing words of a language, are graphemes – in this case <h> and <m>. While these graphemes actually refer to different phonemes, the same procedure yields minimal pairs such as German <fiel> ‘he/she fell’ and <viel> ‘much’ which identifies <f> and <v> as distinct graphemes, even if in these words, they both refer to the same phoneme, /f/. Thus, in the analogical view, graphemes are only parallel to phonemes (and morphemes in morphographic systems, Rogers, 2005, p. 10), but do not (necessarily) stand for phonemes.

Kohrt (1986, pp. 88–89) seriously questions ‘whether it constitutes a reasonable project to transfer discovery procedures that have been designed for a specific substantial domain to a totally different one, imputing that in both areas the problems are just the same’. Indeed, units of speech and units of writing differ in crucial respects, of which Kohrt (1986, p. 89) highlights *segmentation* and *individuation*. The former proves a much easier task for writing than for speech, as it is one of writing’s constitutive features that it is made up of discrete segmental units.¹² This, however, does not mean that all the segments that can be readily identified in writing are automatically graphemes, as per definition, only lexically distinctive units are considered graphemes (but see Lindqvist, 2001 for a different approach). This introduces the problem of *individuation*, or the question of ‘which units that are superficially different should nevertheless be regarded as instances of only one basic unit’ (Kohrt, 1986, p. 89). To answer this question, we need a solid concept of *allography*, which, due to the limited scope of this paper, is proposed in detail elsewhere (Meletis, 2019). The bottom line is that there are two major types of allography, *graphetic* and *graphematic allography*. *Graphetic allography* hinges on the criterion of visual similarity and is, in this vein, conceptually similar to allophony. Concrete graphs such as the three |a|-graphs in <banana> (whether materialised in an individual’s handwriting or in a given typeface such as here) are categorised as graphetic allographs of a basic shape. On the other hand, *graphematic allography* describes, at a higher level, how basic shapes are associated with graphemes. In order to be allographs of a grapheme, basic shapes do not need to be visually similar, but must have the same function. Accordingly, graphematic allography can be compared to allomorphy, where allomorphs can exhibit phonetic similarity but do not have to. An example for graphematic allography is <σ> and <ç> in Greek, where the first occurs word-initially and word-medially and the second appears word-finally. These two units are not separate graphemes since they always

represent the same linguistic unit – in this case the phoneme /s/ – and because, due to their positional distribution, there can logically be no minimal pairs between them.

It is important to note that the analogical view does not negate the referential view but rather represents an analysis at a deeper level. It is a logical precursor to the referential view. Evidence for this comes from the assumption of so-called *grapheme-phoneme-correspondences* in the context of the analogical view. While graphemes are gathered analogously to phonemes – as in <f> and <v> through <fiel> and <viel> –, in the next step, relations between graphemes and phonemes can be explored. The grapheme-phoneme-correspondences that result from this are precisely what the referential view considers to be the minimal units of writing, i.e. graphemes. That both concepts can be reconciled becomes evident in Rezec’s (2009) model, in which both analogical and referential graphemes are units: at a lower level, Rezec assumes a grapheme in line with the analogical view – e.g. the graphemes <n> and <g> – while at a higher level, he postulates the so-called *phoneme image* (German *Phonemabbild*) as a separate unit which is constituted by a grapheme or a combination of graphemes that represent(s) a phoneme, e.g. <ng> which, in many systems, can represent /ŋ/ (Figure 2). Rezec’s model differs from the traditional analogical view in that the grapheme-phoneme-correspondence is treated as a separate *unit* which implies that, contrary to what Günther claims, relations can indeed also be units. A number of approaches in the past have attempted to account for both conceptions within one model. In these models, the analogical grapheme has also been referred to as, among other things, *grapho-grapheme* (Heller, 1980) or *graphemic grapheme* (Herrick, 1994), while the referential grapheme, Rezec’s *phoneme image*, has been termed *phono-grapheme* (Heller, 1980) or *phonological-fit grapheme* (Herrick, 1994).

3. Redefining the grapheme

Both the referential and the analogical definitions of the grapheme only cater to alphabetic – and, though not explicitly, other segmentally phonographic (Gnanadesikan, 2017) – writing systems by relying on the phoneme, a unit that, however, is not at the centre of all types of writing systems. This motivates the present attempt of defining the grapheme cross-grapholinguistically. Firstly, if we take writing as the starting point of analysis – and I strongly believe a description of writing should do that – it is a simple fact that in every writing system, regardless of how much it differs from other systems, there is a minimal unit. This does not imply that it is a trivial task to

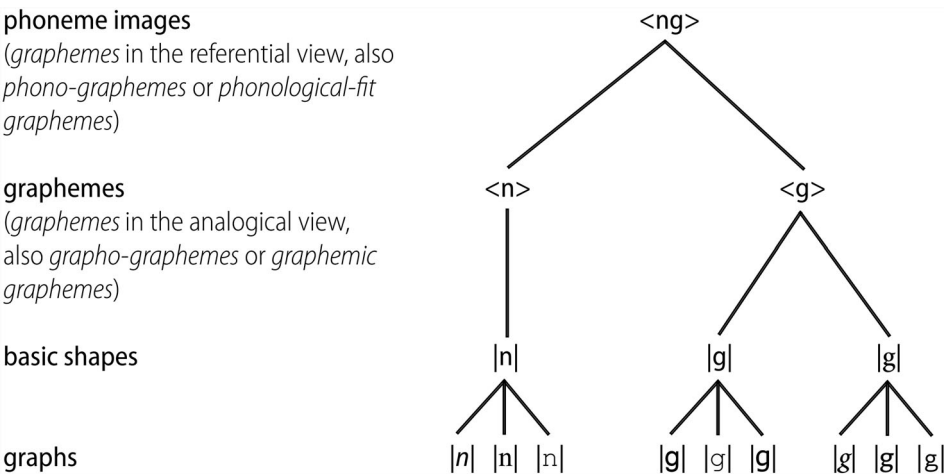


Figure 2. Four-level model of writing systems, adapted from Rezec (2013, p. 244).

decide what this minimal unit is. All writing has by definition at its core the function of encoding language. This alone is reason enough to study the minimal units of writing and analyse comparatively how they serve this function. Secondly, from a theoretical and, more specifically, linguistic standpoint, positing a unit *grapheme* only for a single type of writing system – the alphabetic type (which Glück, 2016, p. 251 does, for example) – is a ‘restriction that would be hard to imagine in the domain of phonology’ (Birk, 2013, my translation)¹³ and other linguistic domains. It would be untenable to suggest a definition of the phoneme or the morpheme that applies only to one language type or even a specific language. The same should apply to the task of defining the grapheme.¹⁴ However, assuming type-specific concepts of the grapheme appears to be standard procedure in grapholinguistics. Some have not dismissed the idea of a more inclusive and general definition of the grapheme, including Sproat (2000, p. 25, emphasis in original), who writes: ‘I will use the term *grapheme* to denote a basic symbol of a writing system’. Without going into detail about what *grapheme* means for him, it is noteworthy that he chooses to treat it as a concept and term applicable to any writing system.¹⁵

In a view that interprets writing systems as semiotic systems, graphemes are semiotic signs.¹⁶ Graphemes typically constitute the smallest semiotic relations between visual units of scripts (basic shapes) and units of language (phonemes, syllables, morphemes, ...).¹⁷ In a grapheme, the visual unit is the *signifier*, the linguistic unit the *signified* (Fedorova, 2013, p. 50). Graphemes are, not unlike morphemes, simultaneously relations and units. In most cases, the semiotic relation between the constituents of graphemes is arbitrary, making them arbitrary signs. However, in the case of iconicity, more specifically in the presence of remnants of pictography (e.g. in the modern morphographic writing system of Chinese), the relationship can be at least partially motivated. While a semiotic analysis of writing is not unproblematic (Harris, 1994), it offers advantages for description: firstly, it allows separating the visual from the linguistic and, consequently, drawing a necessary distinction between *graphetics*, which studies the materiality of writing (i.e. the visual constituent of graphemes) and *graphematics*, which studies precisely the relationship between the visual and the linguistic (i.e. both constituents of graphemes and their interaction).¹⁸ Secondly, this separation makes possible the treatment of writing as its own system with its own idiosyncratic regularities, which are fundamentally determined by the governing principle of *spatiality* (Harris, 2005) and not necessarily or exclusively by the fact that writing represents language. However, at the same time, it allows an analysis of writing doing the latter: visually representing language and, as such, being a complementary *modality* of a given language whose primary modality is either spoken or signed. Note that, as will become evident below, a treatment of graphemes as signs does not equal an adherence to the *referential view*, as graphemes are not straightforwardly defined as visual representations of all phonemes of a language. By contrast, a semiotic analysis is just an acknowledgement of the fact that writing, in its narrow sense, is defined as the representation of language, and must, therefore, represent linguistic information.

3.1. Three criteria for a universal definition of the grapheme

In the following paragraphs, three criteria will be presented which must be met for a unit of writing to be identified as a grapheme. In the sections that follow, examples from all major types of writing systems (following Daniels’s typology, Daniels, 2017) will be given to test these criteria. This will result in a tentative definition of the grapheme that holds across different types of writing systems. Note that it is a structural definition, and that there exist other possibilities of defining the grapheme.¹⁹

Criterion (1): Lexical distinctiveness. Graphemes differentiate meaning. Following the analogical view, this condition can be tested via minimal pairs assembled with existing words of a language – non-words or pseudowords do not count (but see Lindqvist, 2001 for a different approach). In German, for example, there exist minimal pairs such as <Saum> ‘seam’ and <Baum> ‘tree’, but also pairs like <Schaum> ‘foam’ and <Baum>. The contrast of <s> and is segmental, while

the contrast <sch> (for /ʃ/) vs. is graphetically polysegmental since <sch> consists of three basic shapes: |s|, |c|, and |h|. The question is whether <sch> is also graphematically polysegmental, i.e. whether it is a combination of more than one grapheme. This question will be answered by means of criterion (3).

Criterion (2): Linguistic value. While criterion (1) corresponds to the analogists' conception of *grapheme*, criterion (2) conforms in part to the referentialists' claims: a grapheme represents, as in 'stands for', a linguistic unit, although, against the referentialists' focus on the written representation of phonology, this is by no means limited to phonemes, but also includes syllables and morphemes. The fact that in graphemes, a basic shape must stand for at least one linguistic unit to count as a grapheme does not imply that inversely, every linguistic unit in a language's phonology or morphology must be represented by a basic shape. The analytical direction relevant here, thus, is *basic shape* → *linguistic unit*. Accordingly, I agree with the analogists' critique that assuming graphemes (or allographs) such as German <a>, <ah>, <aa>, etc. for the phoneme /a/ or a polysegmental grapheme <ng> for the phoneme /ŋ/ is superfluous. Conversely, it is not necessary to assume, for example, two graphemes <v₁> and <v₂> in German because <v> can represent two phonemes, /v/ and /f/. In order to identify a unit as a grapheme, it is not necessary for it to refer to only one linguistic unit, and its linguistic reference does not need to be stable. It is only imperative that it has a linguistic value in all contexts in which it is used.

In graphemes which were discovered using the analogical method, i.e. which conform to criterion (1) and are much fewer in number than the referential graphemes, basic shapes also stand for linguistic units; crucially, however, this is not the defining criterion. In fact, in German, the basic shape |c| is, on its own, not the signifier of a grapheme. It does, however, differentiate meaning: consider <denken> 'to think' vs. <decken> 'to cover'. Here, criterion (1) is satisfied (Rezec, 2013, p. 231). Interestingly, though, even analogists claim that |c| is not a grapheme and justify this with its limited distribution: in native²⁰ German words, it never occurs without <k> or <h>. A possible and rather probable explanation for this distribution that the analogists seldom explicate is that |c| alone – again, in native words – does not stand for a phoneme, whereas <ch> /x/ and <ck> /k/ do. While <ck> is not a grapheme since it is syllabically determined much like the above-mentioned <ff>,²¹ <ch> is indeed a grapheme of German. It differentiates meaning and it has a linguistic value. But is it minimal? Why is <ch> a grapheme of German and <ng> is not? This is where minimality comes into play, cf. criterion (3) below.

Where criteria (1) and (2) get fuzzy is when it comes to punctuation. Punctuation signs such as the period or the comma definitely have the potential to distinguish meaning, e.g. when <.> distinguishes a declarative sentence from a question, which is marked by <?>. What they refer to, however, remains rather abstract since it pertains to higher linguistic levels – the syntactic, pragmatic, prosodic levels. Consequently, the description of criterion (2) is not final as yet. A more detailed discussion will have to critically reflect on whether punctuation signs are to be treated as graphemes and what this means for the overall definition of *grapheme*. What must also be considered are types of graphetic variation that have graphematic potential, be it bold or cursive print, underlining, colour, or other types of visual design choices. These have been called graphemes in the past (Gallmann, 1985) and have been granted, analogously to prosodic features in phonology, the status of suprasegmentals (Günther, 1988). It is safe to say that these latter types of graphematic functions are not encompassed by a narrow definition of the grapheme. By contrast, graphetic segments such as punctuation signs, digits, special characters, etc. will definitely have to be accounted for by such a definition. Likely, depending on what exactly a grapheme refers to – whether it is a single, concrete linguistic unit such as a phoneme or a morpheme or instead linguistic information or a linguistic function – it will be imperative to assume different classes of graphemes. I leave this open for future discussion and concentrate on the default type of graphemes, which are those that refer to specific linguistic units.

Criterion (3): Minimality. A crucial problem in discussing the grapheme as a basic unit of writing is the understanding of minimality. Above, it was postulated firmly that the graphemes in any

given writing system should be identified based on the written side of the semiotic relation, not the linguistic side. This implies that what is minimal is, if not completely, then at least to a large degree, determined graphetically – in other words, visually. This underlines that graphematics can never be analysed completely divorced from graphetics. Similar to Lindqvist's (2001) approach, the first step that needs to be undertaken is to gather the basic shapes used in a writing system. Basic shapes themselves are not yet categorised – as letters, digits, punctuation signs, or special characters. Visually, there is no categorical difference between <2> and <Z>, and it is a matter of graphematic convention and competence to know that they are part of different inventories. Thus, at the point of purely graphetic analysis, they are all simply basic shapes, including shapes such as |♥|. The central feature of basic shapes that can be evaluated independently of their graphematic status is their segmentality. A basic shape occupies exactly one segmental space of the writing surface (Meletis, 2015; see Figure 3). As will be shown below, in writing systems in which vowel graphemes are visually less salient since they are smaller in size to consonant graphemes and attaching to them, graphemes can also be subsegmental, i.e. subparts of what occupies a horizontal segmental space can function as a grapheme.

After we have proceeded descriptively and collected all the segmental graphetic material from a product of writing, the next step is to test whether the gathered basic shapes meet criteria (1) and (2). In German, a sequence of two basic shapes such as |ng| is not a grapheme <ng> since numerous minimal pairs can already be found for either constituent: <Bank> 'bank' vs. <bang> 'anxious', <bang> vs. <Band> 'tape', <bang> vs. <Balg> 'brat'. Thus, the fact that the sequence (or *digraph*) <ng> represents the phoneme /ŋ/ and thus meets criterion (2) does not make it a grapheme since criterion (1) is already met independently by its segmental constituents (see Berg, 2019, p. 30, who argues similarly). In this vein, the much-debated and above-mentioned sequence (or *trigraph*) <sch> should not be considered a grapheme of German since minimal pairs can be found for both <s> and <ch>: <Masche> 'bow' vs. <manche> 'some', <Masche> vs. <Maske> 'mask' (Fuhrhop & Peters, 2013, p. 205). The situation is different for <ch> and <qu>. No minimal pairs can be found for their individual segments – at least not both of them. In <qu>, the |q| can be substituted in a very limited range of minimal pairs, including <Quelle> 'source' vs. <Duelle> 'duels'. However, the <u> of <qu> can never be substituted. The distribution of |q| is peculiar, but, strictly speaking, because of the mentioned minimal pairs, it meets criterion (1) and, per the analogical view, should be a grapheme. However, it does not meet criterion (2) as it does not stand for any linguistic unit; thus, |q| alone is not a grapheme. Cases such as <ch> or <qu>, in which one segment is not a grapheme itself – or segments which are combined are not graphemes individually – are conceptually still treated as graphemes, i.e. graphematically segmental units, although graphetically, <ch> and <qu> are polysegmental. To summarise, the criterion of graphematic minimality is met either by a single basic shape for which criteria (1) or (2) apply, or by a sequence of basic shapes in which for at least one of the basic shapes, criteria

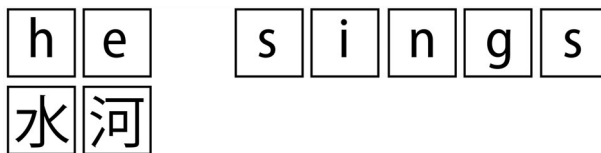


Figure 3. Basic shapes occupy one segmental space, highlighted by black squares.

Note: The blank space between <he> and <sings> is neither a basic shape nor a grapheme. It is not a basic shape since it is literally invisible by consisting of no graphic material whatsoever. It is only made visible by its graphic context, i.e. non-blank material surrounding it. As for it being a grapheme: in conceptions in which null-graphemes are assumed, the blank space is lexically distinctive, as in <hand> vs. <and>, where <h> contrasts with Ø. However, it never has linguistic value, i.e. represents no linguistic unit, and is thus, in the present conception, not regarded as a grapheme. Also, as an anonymous reviewer correctly noted, the second basic shape from Chinese script, [河], is complexly structured as it consists of two components. However, as elaborated in Section 3.2., these subsegmental components which originate from independent graphemes together form a new and single basic shape and take up one segmental space.

(1) and/or (2) do not apply, as for <ch>. These latter sequences are not graphetically minimal, but they are graphematically minimal.

One problem that remains in the definition of the grapheme that concerns the criterion of minimality is subsegmentality (Lindqvist, 2001, p. 10). If the grapheme is supposed to be the minimal contrastive unit, then why is not the hasta²² (or head) || (or its position) in |b| vs. |p| a grapheme, or the lowest horizontal stroke of the |E| in |E| vs. |F|? For these cases, the answer is fairly simple: Neither the hasta (or its location) in |b| and |p| nor the lowest stroke in |E| refer to linguistic units or have linguistic value (but cf. Primus, 2004), i.e. they do not meet criterion (2). The question of subsegmentality is a lot more complex in Chinese, however, which will be discussed next.

3.2. Graphemes in morphographic Chinese

As a morphographic writing system, Chinese is often contrasted with alphabets and positioned at the other extreme of the typological spectrum. Unsurprisingly, its basic units are seldom compared with the basic units of phonographic writing systems. However, as the proposed definition of *grapheme* should be applicable to every writing system, the three suggested criteria need to be tested for Chinese as well.

Consider the two units <請> *qǐng* 'please, to ask' and <情> *qíng* 'emotion'.²³ They are complex in that they are constituted by two components: a semantic component (or *radical*), which, in this case, is positioned on the left, and a phonological component (or *phonetic*), located on the right (Myers, 2019, Chapter 2). They share the same phonological component, which indicates the pronunciation as it is derived from <青> *qīng* 'green/blue' except for tone. Their meaning is differentiated by the differing semantic components, which, therefore, meet the criterion of lexical distinctiveness.²⁴ Seen as the complexly structured <請> and <情> take up only one segmental space each, according to the analogical view, subsegmental components – semantic, but also phonological (for which there are also minimal pairs, see below) – would have to be accepted as graphemes. This is indeed what both DeFrancis (1989) and Sproat (2000) suggest. It constitutes the crucial difference between a polysegmental sequence of two graphemes in an alphabet such as <ng> and the graphematically segmental <請> *qǐng* 'please, to ask': whereas in <ng>, both <n> and <g> retain their status as graphetic segments and both have the potential to refer to a linguistic unit (a phoneme) individually, in the Chinese example, two characters which are graphemes individually – <言> *yán* 'speech' and <青> *qīng* 'green/blue' – are 'cramped' together into one segmental space. The question is, now, whether they still individually meet the linguistic value criterion when they, as minimised versions of their original graphemes, are part of this new, complex structure.

The semantic components in these examples indicate meaning, although they do so in a fairly abstract way, by signalling an approximate semantic clue. The phonological components, different in their functionality, signal the pronunciation of a character. Prototypical graphemes in Chinese stand for morphemes, and neither of these components does that: in <請> *qǐng*, the <言> *yán*-part, i.e. the speech-radical, refers to one facet of the meaning of the morpheme ('speech'), but it is not in every complex grapheme that the semantic component straightforwardly points to the right meaning of the morpheme; it can also be opaque in this respect (Ho, Ng, & Ng, 2003). Examples are <河> *hé* 'river' and <汗> *hàn* 'sweat'. In this pair, the water-radical <氵> only indicates semantic affiliation, but the specific signified of the grapheme <水> *shuǐ* 'water' from which the radical derives is blurred in the process.²⁵

The other component in <請> *qǐng*, <青> *qīng*, is used as a phonological component, i.e. only for its phonological representation. It contributes a part of the form, i.e. the signifier, of the morpheme it originally represented. However, for the reader, recognising the morpheme that the original grapheme refers to is necessary to access its phonological representation, even if the meaning of that morpheme is completely disregarded in the final complex grapheme, e.g. 'green/blue' in

<請> *qǐng* ‘please, to ask’. No graphemes or components of graphemes in Chinese are directly phonographic.

Like <ng>, which conforms to the linguistic value criterion by referring to the phoneme /ŋ/, the graphemes <請> *qǐng* and <河> *hé* represent morphemes on their own. However, their subsegmental components, in their minimised sizes, cannot stand alone in the same way that <n> and <g> can. It might be tempting to argue that in a grapheme such as <請> *qǐng* ‘please, to ask’, the semantic component represents the signified of the morpheme, while the phonological component represents the signifier of the morpheme. However, as shown, it is not as simple as that: semantic components often only give a vague clue about the meaning of a morpheme, and due to language change, the clues to the pronunciation that the phonological components give have also frequently become opaque (Qiu, 2000, pp. 20–21, 247–252; Sampson, 1994), so that now only about 19 % of phonological components accurately predict the phonological representation of the morpheme represented by the grapheme (Ho & Bryant, 1997, p. 279).

In this context, consider also the *kanji*-part of the Japanese writing system: *kanji* are morphographic graphemes, many of which are derived from Chinese graphemes. These loaned graphemes often include phonological components. However, since many *kanji* represent native Japanese morphemes, and these morphemes exhibit Japanese phonological representations, there is no link between originally Chinese phonological components and the native Japanese pronunciation of morphemes (cf. Sampson, 1994, p. 128). These *kanji* are devoid of any phonography, i.e. they are purely morphographic graphemes.²⁶

In sum, I argue that characters of the type <請> *qǐng* and <河> *hé* are complex graphemes rather than sequences of two graphemes because the two initially independent graphemes that are shrunk in size and fit together into one segmental space (and in this minimised size, cannot occur alone) change their function and cease to meet the linguistic value criterion: they no longer stand for morphemes, and this distinguishes the subsegmental components fundamentally from their individual segmental counterparts, which are undoubtedly graphemes.²⁷

What these examples highlight is that there exist features of writing that cannot be explained by phonology or other linguistic levels: the inner systematics of the type of complex graphemes in Chinese characterised above – with one component signalling meaning, the other pronunciation²⁸ – are a genuine graphematic feature that has no parallel in any other linguistic level. Writing represents language, but that does not mean that all of writing’s features should or even can be explained by this representational function.

3.3. Graphemes in non-alphabetic phonographic writing systems

Up to this point, the three criteria of the proposed definition of *grapheme* have been illustrated with examples from an alphabet – German – and a morphographic writing system – Chinese. In principle, this already underlines their applicability to other alphabets and morphographic writing systems. This section will now deal with non-alphabetic phonographic writing systems and some of the challenges they pose for the proposed conception of the grapheme.

As mentioned above, *spatiality* is the governing principle of the written modality of language. Thus far, writing has been treated predominantly as one-dimensionally (most frequently horizontally) linear, a situation referred to as *dogma of linearity* (Krämer, 2003, p. 159). For this reason, mainstream linguistics largely excluded any considerations of two-dimensional features of writing (Waller, 1991, pp. 354–357), leading to the misconception that – in horizontal writing systems – graphemes only appear next to each other and must occupy a relatively equal amount of space. Given the examples from German and Chinese above, it indeed appears as if only basic shapes, i.e. graphetic segments which occupy their own segmental space, can function as graphemes and that they do so only holistically, implying that there are no subsegmental graphemes. However, both *abjads*, in which, by default, only consonant phonemes are graphematically represented, and many *abugidas*, in which default graphemes represent

consonant phonemes (C) and a vowel phoneme (most often /a/) which is inherent in the grapheme, so that /Ca/ is represented by <C>, appear to contradict this claim. They demonstrate that two basic shapes which share a segmental space on the horizontal axis can be separated on the vertical axis.

Both abjads and abugidas exhibit vowel graphemes which, on the horizontal axis, are graphetically subsegmental: in abjad²⁹ Arabic, they are optional and rarely used, in abugidic Thai, they are obligatory. Arabic <ﺭ> /r/ is a combination of the consonant grapheme <ﺭ> /r/ and the (in most contexts optional) vowel grapheme <ْ> /a/ which share one segmental space horizontally but not vertically (see Figure 4). The writing system of Thai functions similarly: in a vertical sequence such as <ดี> /di:/ 'good', the two graphemes representing the phonemes /d/ and /i:/, respectively, share one horizontal segmental space. Vertically, however, the basic shape |[◌]| which materialises the vowel grapheme /i:/ is positioned in a separate space on top of the primary consonant grapheme <ด>. Note that |[◌]| always materialises /i:/ and can be systematically combined with every consonant grapheme, making it a vowel grapheme <◌[◌]>. The fact that the vowel graphemes in these two examples do not occupy their own segmental space horizontally reflects the fact that they are dependent, i.e. cannot occur on their own, at least not in post-consonantal position (Rogers, 2005, p. 11 calls them *bound graphemes*); however, in some writing systems, they have corresponding allographs that occur initially and that are independent (see below).

Aside from the necessary inclusion of the vertical axis, abugidas raise a number of additional open questions for the proposed definition of *grapheme*. Their main unit is the so-called *akṣara*, which is why they are commonly also referred to as *akṣarik systems* or *akṣara-based systems* (Gnadesikan, 2017, p. 19). Akṣaras correspond neither straightforwardly with phonological syllables nor with phonemes. Instead, they are subsyllabic units whose central component is a long or short vowel that may be preceded but not followed by a consonant or consonant cluster (Patel, 2010, p. 3): V, CV, CCV, CCCV, etc. (Salomon, 2007, p. 28). In syllables, by contrast, vowels can also be followed by consonants. As in the Thai example above, vowel graphemes can be smaller in size than consonant graphemes and bound, i.e. dependent on them, and are positioned above or below them. However, vowel graphemes can also be of equal size and occupy their own segmental space on the horizontal axis, such as Thai <า> which represents /a:/ and occurs to the right of the consonant grapheme as in <ภาษา> /p^ha:sa:/ 'language'. Also in Tamil, another akṣara-based system, there are dependent vowel graphemes that are equal in size to the independent consonant graphemes and appear in-line with them (Bhuvaneshwari & Padakannaya, 2014, p. 192).

As mentioned, many akṣara-based systems have an additional set of allographs of vowel graphemes that occur initially and are independent and of the same size as consonant graphemes: an example is Devanāgarī <उ> which represents /u/ in initial position as opposed to <उ> which represents the same vowel phoneme post-consonantly. This raises the question whether these two variants are two separate graphemes or whether they are allographs (see also Section 2.4.). Here, the answer is clear: they are positional graphematic allographs similar to <σ> and <ς> in Greek which both represent /s/ in different positions. Due to their complementary distribution there exist no minimal pairs between them and they always represent the same linguistic unit.

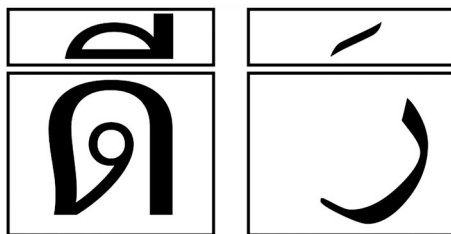


Figure 4. Separate vertical spaces in Thai and Arabic.

In some akṣara-based systems, vowel phonemes are represented by a combination of two discontinuous basic shapes, i.e. basic shapes that do not appear next to each other, but are separated by a consonant grapheme between them (indicated by a dash), a construction reminiscent of circumfixes in morphology. Examples come from Tamil as in <௫—ஈ> which represents /o/ or Thai <โ—ฮ> which represents /e/. The question, now, is whether these graphematic representations of vowels are single graphemes or not. In these two examples, the answer is no, and the reason is the criterion of minimality. In Tamil, both constituents are already graphemes individually: <௫> represents /e/ and <ஈ> represents /a:/.³⁰ The same is true for the Thai example, where <โ> stands for /e:/ and <ฮ> for /a/.³¹ Even if these combinations together create a new graphematic value that is not the compositional sum of its parts' functions, they are not complex graphemes but sequences of two separate graphemes, respectively. That way, these examples echo German <sch> which represents /ʃ/ but is not a grapheme since <s> /s/ and <ch> /x/ are already graphemes themselves. Note that if the two units in such spatially discontinuous combinations were not already both individual graphemes or if only one of them was a grapheme (as in <ch>, where only <h> is a grapheme, which makes <ch> a complex grapheme, cf. Section 3.1.), such combinations of non-adjacent basic shapes can also constitute graphemes.

Another common feature of these systems is that akṣaras often start with consonant clusters, i.e. consonants without a vowel between them, and the basic shapes that materialise these conjunct consonant graphemes are frequently contracted to ligatures. In Devanāgarī as used in Hindi, for example, <घ> /gʱ/ and <र> /r/ are contracted to <घर> /gʱr/. Graphetically, this is reminiscent of the above-mentioned complex graphemes of Chinese in which individual characters that each occupy their own segmental space change their shape – most importantly, become smaller – to fit into one segmental space when they are repurposed as semantic or phonological components. The same happens in <घर>. The difference between this example and the Chinese examples is that even though the Devanāgarī basic shapes are contracted and are no longer graphetically segmental (and so intertwined that a separation is almost impossible), they still retain the function they had when they were individual shapes – <घ> and <र> –, meaning the ligature still represents two individual linguistic units (phonemes) and satisfies the linguistic value criterion, i.e. it constitutes two (subsegmental) graphemes. The same is the case for irregular, non-standard combinations of consonant and vowel graphemes such as they occur, for example, in Tamil.³² In sum, these ligatures, which are common in akṣara-based systems, are not single complex graphemes, but sequences of two or more individual graphemes that are graphetically segmental.³³

Korean – despite its 'featural' features (Sampson, 2015) typologically most reasonably characterised as an alphabet³⁴ – is an extreme example in this respect: here, graphemes, referring to phonemes, are graphetically subsegmental as they are complexly arranged in syllable blocks which are themselves graphetically segmental. The segmental space is thus filled with graphematic clusters which refer to phonological syllables. For example, Korean <ㄱ> stands for /k/ and meets all the criteria of the proposed definition of *grapheme*, but it is not used linearly and usually does not occupy its own segmental space. Instead, it is combined with other graphemes as in <ㄱㅏ>, which stands for the syllable /kak/. This graphematic syllable block, although it is graphetically segmental,³⁵ is no grapheme, but a combination of three graphemes. As the subsegmental shapes unambiguously refer to linguistic units (phonemes), graphemes can definitely be graphetically subsegmental, as long as they still meet all of the grapheme criteria. This makes them distinct from the subsegmental components in Chinese which work only compositionally (see above). However, it cannot be denied that in Chinese, there is still a relevant graphematic level that is subsegmental, as obviously, semantic and phonological components, elements smaller than graphetic segments, i.e. basic shapes, can potentially have graphematic value. This is similar to what Primus (2004, 2006) proposes for the German writing system or more generally, writing systems using Roman script, observing that parts of basic shapes are graphematically related with phonological features. Other examples of subsegmental graphematic functions include the [̥]-diacritic in umlaut graphemes such as <ö> and <ü> in German which graphematically marks a change of the

|³⁶ | which, when it is added to Hiragana graphemes, marks voicing of the initial consonant phoneme in the syllables represented by the graphemes: <き> /ki/ becomes <ぎ> /gi/.³⁶

Note that this section is incomplete. It is to be expected that writing systems not treated here provide additional open questions for the proposed definition of *grapheme*. These will have to be addressed when they are encountered. However, the fact that the above-mentioned idiosyncratic features of various systems can be explained within the proposed conception points to the probability that the true core features of the grapheme have been identified.

Table 1. Criteria for the definition of *grapheme* applied to units from different writing systems; graphemes are highlighted in grey.

	criterion (1) <i>lexical distinctiveness</i>	criterion (2) <i>linguistic value</i>	criterion (3) <i>minimality</i>
German <ng>	+	+ phoneme /ŋ/	– both <n> and <g> are already graphemes
German <ch>	+	+ phoneme /ç/	+
German <sch>	+	+ phoneme /ʃ/	– both <s> and <ch> are already graphemes
German c <den <u>ck</u> t> vs. <de <u>ck</u> t>	+	–	+
German (hasta) in <p> vs. 	+	–	+
Chinese <河> 'river'	+	+ morpheme {river}	+
Chinese ǎ	+	– semantic component 'water', which itself represents no linguistic unit	+
Thai <ด>	+	+ phoneme /d/	+
Thai <ิ>	+	+ phoneme /i/	+
Korean <ㄱ>	+	+ phoneme /k/	+
Korean <ㄱㅏ>	+	+ syllable /kak/	– is made up of three graphemes that refer to phonemes, respectively
Japanese <き>	+	+ mora /ki/	+
Arabic <ر>	+	+ phoneme /r/	+
Arabic <ا>	+	+ phoneme /a/	+
Arabic <را>	+	+ CV cluster/syllable /ra/	– both <ر> and <ا> are already graphemes
Tamil <ஒ—ர>	+	+ phoneme /o/	– both <ஒ> and <ர> are already graphemes
Hindī (Devanāgarī) <घ्र>	+	+ consonant cluster /gʱr/	– both <घ> /gʱ/ and <र> /r/ are already graphemes

4. Conclusion

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. Table 1 collects the cited examples from various types of writing systems and shows the three criteria at work, highlighting that this definition is applicable cross-graphologically.

However, this proposal for a more universal conception of the grapheme is precisely that: a *proposal*, and I do not claim that it is the (only) right one. It is tentative, since it will, of course, have to be tested for even more writing systems. The description of grapheme inventories in the world's writing systems requires, aside from a definition of *grapheme*, a solid accompanying concept of *allography* (Meletis, 2019), and idiosyncratic features of writing systems that challenge these conceptions of the grapheme and allography need to be addressed in more detailed case studies.

Like Berg (2019, p. 32–33), I want to generally advocate for more calmness and tolerance in the sometimes heated discussion of a grapheme definition. It does not make sense to speak of 'one grapheme inventory' of a given writing system, as different conceptions arrive at different – and possibly equally justified – inventories. As different analyses are driven by distinct epistemological interests, it would be unfair to proclaim a priori that only one method is the right one.

As a descriptive comparative concept (Haspelmath, 2010), the grapheme makes possible a unified description and, in turn, comparison of typologically highly diverse writing systems. It does not negate the fact that the specifics across systems differ, as it only claims that the core is always the same. This is unsurprising, since writing, no matter how drastically different various systems might appear, always encodes language. The next and necessary step is to evaluate this conception of the grapheme in the light of available data from the processing of writing. Are graphemes as defined here psychologically real? What role do they play in the acquisition of reading and writing processes and are they relevant once these processes are mastered? It is crucial to note, however, that whatever the answers to these questions might be, they only partially affect the value of the grapheme as a descriptive concept. Descriptively, it opens the door for a graphology that looks beyond typological boundaries and strives for a theory of writing that encompasses all kinds of writing systems.

Notes

1. In this context, Mugdan (1990, p. 50, emphasis in original) comments on the problematic vagueness of blankly coining linguistic terms ending in *-eme*: 'The insatiable need for designations of linguistic units ensured the remarkable success of the new derivational pattern, but the combinations of *-eme* with suitable roots (*morph-*, *graph-* and the like) have such general meanings ("unit of form", "unit of writing" etc.) that one could utilize them to name any of several different concepts'.
2. But see, for example, Häffner's (2009) treatment of the German and Japanese writing systems within one theoretical framework.
3. Common consensus is that the term *grapheme* was coined independently at least twice: in addition to Bau-douin de Courtenay, Kohrt (1986, pp. 82–83) mentions Finnish linguist Aarni Penttilä as having introduced the term – albeit in a different sense – in 1932. Furthermore, Daniels (2018, p. 171) claims that in the 1930s, Benjamin Lee Whorf, W. Freeman Twaddell, and R. H. Stetson also all independently used the term *grapheme*.
4. In this view, speech and writing are seen as two modalities of an abstract amodal language system; they both materialise language. Speech and, crucially, also sign language, are primary modalities of such an amodal language system, which means one of them is obligatory in every language. Writing, by contrast, is an optional modality: there are many languages without writing, but there is no language that has neither speech nor sign language as a primary modality.
5. Daniels (2018, p. 168) also argues that there is no *duality of patterning* (or *double articulation*) in writing, but see Holenstein (1983) and Ladd (2014, Chapter 5.4.2) for a different view.
6. The *basic shape* (from German *Grundform*, Rezec, 2009, 2013) is a material yet abstract unit. It represents a 'skeleton', a bundle of visual features that are necessary to perceptually distinguish a shape from the other shapes in an inventory (Herrick, 1974, p. 11), which include the number of elements (straight lines, curves, dots, etc.), the position of these elements in space and the relation between these elements (connections

and types of connections). The basic shape is itself the sum of all *graphs*, which are etic and concrete substantiations of basic shapes: |a|, |a|, and |a| are all (graphetic allo-)graphs of the basic shape |a|, but |a| and |a| are not, as they are assigned to a different basic shape, |a| (cf. Section 3.1.).

An anonymous reviewer asked how the conception of *grapheme* proposed here can accommodate the infinite visual variety of graphs, i.e. concrete realisations of basic shapes in handwriting but also in print (considering hundreds of thousands of existing typefaces). The assignment of graphs to basic shapes is exclusively a visual, i.e. graphetic matter and does not affect the assignment of graphs to graphemes. Variation between graphs of the same basic shape, e.g. |a| and |a|, thus, is always non-graphematic, i.e. denotatively non-distinctive (it can have different connotations, however). Only if graphs become so visually dissimilar that they are assigned to different basic shapes which have distinct graphematic values in a writing system do visual differences become important: consider the respective relations of the horizontal and vertical strokes in |T| and |Γ|, two basic shapes of Greek script which, in the Greek writing system, embody different graphemes.

7. A *script* is defined here as an inventory of basic shapes without an association to a given writing system or language (Coulmas, 1996, p. 1380). Many writing systems, for example English, German, Finnish, etc., employ Roman script, many others like Russian, Ukrainian, etc. employ Cyrillic script, etc.
8. Graphetic units are enclosed in vertical strokes | | and graphematic units in angle brackets < > (Berg & Evertz, 2018, p. 190).
9. The criterion itself might be generalisable across modalities, but what is salient is specific to each modality.
10. Schmidt (2018, p. 128) points out a problem of treating skeletal positions as graphemes: in graphematic words such as <beten> ‘to pray’ or <lesen> ‘to read’, the first vowel, in each case an instance of <e>, is associated with two skeletal positions, respectively, since these graphematic words are simultaneously, as graphematic feet, trochees, i.e. their first graphematic syllables – <be> and <le> – are ‘strong’ or ‘prominent’. As Schmidt argues, it would be absurd to claim that the first <e> in these words represents two graphemes because it is associated with two skeletal positions. Skeletal positions, thus, do not straightforwardly correspond with graphemes. Schmidt adds that Primus and colleagues silently acknowledged this by changing the designation of the skeletal positions from “G” (for grapheme) to the more neutral “X” in newer versions of the model (cf. Figure 1).
11. Henderson (1985) describes the uses of the term *grapheme* in the realm of psychology and calls the use that essentially corresponds to the analogical view *Sense 1* and the use that roughly corresponds to the referential view *Sense 2*.
12. This often does not hold for cursive handwriting, where the basic shapes are connected to each other, as well as in some scripts such as Arabic, where the basic shapes are mostly connected. Even if they are visually connected, however, the units are still learnt and stored as separate units, as segments.
13. ‚[...] eine Einschränkung, wie sie im Bereich der Phonologie nur schwer vorstellbar ist‘ (Birk, 2013).
14. However, the picture might not be quite so simple. A discussion initiated by Haspelmath (2010, p. 663) focuses on the distinction between ‘descriptive categories, that is, categories of particular languages, and comparative concepts, which are used for crosslinguistic comparison and are specifically created by typologists for the purposes of comparison’. The grapheme might be more reasonably conceived of as a more broadly defined comparative concept, which would allow for specific graphemes of different writing systems to differ in some respects. This also opens up the theoretical possibility for different concepts of *grapheme* that are specific to given types of writing systems, such as the phonographic grapheme vs. the morphographic grapheme. However, what I want to categorically reject is the idea that there can be a grapheme in one type of writing systems – such as alphabets – but not in others – such as morphographic writing systems. If we assume graphemes, there are graphemes in every (type of) writing system, even if their specific features differ.
15. He adds, however, that ‘in discussing some writing systems we may use the term grapheme in slightly different ways depending upon how fine-grained an analysis is being assumed’ (Sproat, 2000, p. 25). This reflects that even though Sproat intends to use the same term for all writing systems, he is aware of the differences it entails for different writing systems.
16. A number of recent approaches (Fedorova, 2013; Klinkenberg & Polis, 2018; Rizza, 2018) opt for such a semiotic analysis. Note that by choosing this analysis, I in no way claim to give an answer to the complex question of *how* written units signify units of language (Harris, 1994, p. 45), as I merely posit that written units *do* signify units of language.
17. This is also in accordance with Weingarten (2011, p. 17), who defines *writing system* as ‘an ordered pair of a single language and a single script: (Language_L, Script_S)’, for example (German_L, Roman_S) or (Russian_L, Cyrillic_S). The writing systems are always named after the partaking languages: e.g. German writing system, Russian writing system.
18. An anonymous reviewer asked whether the division in *graphematics* and *graphetics* is not simply Gelb’s (1969) division of *inner* vs. *outer* form. While the two certainly correspond with each other, the former is in some way a systematisation of the latter and, more importantly, an integration into a conceptual and terminological linguistic framework. It takes the fact that writing has an inner, functional side – graphematics – and an outer,

formal side – graphetics – further and allows for both to be studied in detail by eponymous grapholinguistic subdisciplines.

19. One of these possibilities is based on conventionality (Valkama, 2017). In this vein, for example, an anonymous reviewer noted that in Spanish, <ll> is unequivocally considered to be a grapheme, which is reflected in standard Spanish lexicographic practice. This raises the question of how the community of users of a writing system evaluates a given unit. While this question is interesting and valuable, it is not a structural question, but a question of conventionality. Given the structural criteria presented here, <ll> is not a grapheme since it is not minimal, as <l> is already a grapheme.
20. It does occur alone in loanwords such as <Clown> ‘clown’ or <Cello> ‘cello’. For the graphematic relevance of the assumption of different classes of vocabulary in a language, e.g. native vs. foreign words, cf. Neef (2015, pp. 717–718).
21. This way, <ck> is the alternative of a doubling of <k>, <kk>, with |c| marking the end of the first graphematic syllable and <k> marking the beginning of the second graphematic syllable, cf. German <blic.ken> ‘to look’ (Fuhrhop & Peters, 2013, p. 227; Fuhrhop & Schmidt, 2014, p. 557).
22. Fuhrhop & Buchmann (2009) segment the minuscule basic shapes of Roman script and conclude that all of them consist of a *hasta* (from Latin *hasta* ‘spear’) – in most cases the vertical stroke first produced when writing a basic shape – and a *coda*, a finding going back to the work of Brekle (1995, p. 5) and his *hasta-coda-principle*. In |d|, for example, the *hasta* is the vertical stroke || and the *coda* is the curve |c|.
23. The provided pronunciations are from the Mandarin variety of Chinese.
24. However, these components only form minimal pairs with other subsegmental components. <請> *qīng* ‘please, to ask’ and <情> *qíng* ‘emotion’ are a minimal pair, but <請> *qīng* ‘please, to ask’ vs. <言> *yán* ‘speech’ are not.
25. In this example, it is also evident that the initial basic shape, if used as a subsegmental component, significantly changes its form. This visualises the shift in identity from an independent, segmental grapheme to a subsegmental part of a grapheme.
26. Of course, as morphographic graphemes, they represent morphemes, and morphemes have phonological representations. This phonological representation, however, is in no way graphematically marked in purely morphographic graphemes.
27. Thus, Daniels’s (2018, p. 168) objections that ‘there would be two entirely different kinds of grapheme in the writing system, the semantic ones and the phonetic ones, and they would only achieve specific reference in combination’ and that ‘this also leaves the residue of non-composed characters to get some sort of separate description’ can be relativised. Not every unit that has a graphematic function has to be a grapheme. The fact that ‘composed’ graphemes in Chinese (such as <請> *qīng*) are complex is something we need to be aware of in a graphematic analysis, but it does not mean that we have to treat them differently than non-composed graphemes (such as <言> *yán* and <青> *qīng*). Both of them differentiate meaning, both of them refer to linguistic units (morphemes), and – as elaborated – both of them meet these specific criteria in a minimal manner.
28. There are other types of complex graphemes as well, e.g. graphemes that consist of two or more semantic components (for a typology, see Qiu, 2000, Chapter 6).
29. Note that in an abjad, by definition, only consonant phonemes are graphematically represented. However, in Arabic, long vowels are graphematically represented, and short vowels can optionally be written, too. If the optional short vowel graphemes are produced – as in the given examples – both vowel and consonant phonemes are consistently graphematically represented and Arabic thus functions more like an alphabet, rendering the discussion of these examples in a section on non-alphabetic writing systems a bit misleading. However, given their size, placement, and dependence, (short) vowel graphemes are not equal in status to consonant graphemes even in vowelised Arabic. This significantly distinguishes vowelised Arabic from alphabets and warrants the discussion of its vowel graphemes together with vowel graphemes of abugidas with which they share many features.
30. In this context, an anonymous reviewer brought up the initial and independent vowel grapheme <ஒ> which, just like the discontinuous grapheme combination <ஒ—ஈ>, represents the phoneme /o/. In this case, given the present conception of the grapheme, the initial vowel grapheme does not have a dependent *allograph* like other initial vowel graphemes in Tamil do since <ஒ—ஈ> are two graphemes. However, as this example shows, combinations of graphemes can have non-compositional linguistic values and play important roles in the graphematic representation of the linguistic units (here: phonemes) of a given language. By definition, <ஒ> and <ஒ—ஈ> are no allographs, but they are graphematic representations of the same vowel phoneme – one being a single grapheme and the other a combination of graphemes.
31. <ஐ>, however, is also used to graphematically mark the shortness of vowels, which is obviously the function it fulfils in this combination.
32. Of course, graphetically non-segmentable clusters of graphemes such as the non-standard combination <நு> /nu/ in Tamil are a challenge for the present proposal of a concept of *grapheme*. Instead of analysing them as sequences of consonant and vowel graphemes which can visually not be segmented, one could also admit

the possibility that aside from consonant graphemes and vowel graphemes, there are additionally also syllabic graphemes in Tamil (and similar akṣara-based systems) which refer holistically to phonological syllables. The pros and cons of such an interpretation will need to be discussed in more detail in a specific graphematic analysis of Tamil (or similar systems) that tests the definition of *grapheme* as proposed here.

33. Rogers (2005, p. 12) calls them *non-structural ligatures* and additionally mentions the typographically motivated ligature in Roman script, where and also retain their status as graphemes.
34. Gnanadesikan (2017, p. 29) calls it a *fully vowelised syllabically arranged featural segmentary*, i.e. regards the syllabical arrangement and the graphemes' partially pictographic representation of places of articulation as features of a phonographically segmental writing system, an alphabet.
35. The segmental syllable blocks of Korean could, of course, also be interpreted as complex arrangements of smaller segmental spaces which are occupied by graphemes. Due to the many spatial possibilities of combining these graphemes, however, it appears theoretically more economical to interpret the syllable blocks as graphetically segmental and the graphemes as subsegmental.
36. These examples imply that there are limits to a segmental graphematic analysis that is based on segmental minimal pairs. As it advances, graphematics (and grapholinguistics in general) will likely evolve in similar ways to phonology and develop more fine-grained featural analyses such as the ones already partially proposed by Primus (2004, 2006). Again, this present proposal is just a starting point.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Dimitrios Meletis  <http://orcid.org/0000-0002-8889-6459>

References

- Berg, K. (2019). *Die Graphematik der Morpheme im Deutschen und Englischen* [The graphematics of morphemes in German and English]. Berlin: De Gruyter. doi:10.1515/9783110604856
- Berg, K., & Evertz, M. (2018). Graphematik – die Beziehung zwischen Sprache und Schrift [Graphematics – the relation between language and writing]. In S. Dipper, R. Klabunde, & W. Mihatsch (Eds.), *Linguistik* [Linguistics] (pp. 187–195). Dordrecht: Springer. doi:10.1007/978-3-662-55589-7_10
- Berg, K., Primus, B., & Wagner, L. (2016). Buchstabenmerkmal, Buchstabe, Graphem [Letter feature, letter, grapheme]. In U. Domahs, & B. Primus (Eds.), *Handbuch Laut, Gebärde, Buchstabe* [Handbook sound, sign, letter] (pp. 337–355). Berlin: De Gruyter. doi:10.1515/9783110295993-019
- Bhuvaneshwari, B., & Padakannaya, P. (2014). Reading in Tamil: A more alphabetic and less syllabic akṣara-based orthography. In H. Winkler, & P. Padakannaya (Eds.), *South and Southeast Asian psycholinguistics* (pp. 192–201). Cambridge: Cambridge University Press.
- Birk, E. (2013). Graphem [Grapheme]. In M. Neef, S. Sahel, & R. Weingarten (Eds.), *Schriftlinguistik* [Grapholinguistics]. Berlin: De Gruyter. Retrieved from https://www.degruyter.com/view/WSK/wsk_id_wsk_artikel_artikel_14530
- Brekke, H. (1995). Neues über Groß- und Kleinbuchstaben. Theoretische Begründung der Entwicklung der römischen Majuskelformen zur Minuskelschrift [Something new about uppercase and lowercase letters. Theoretical reasons for the development of Roman majuscule forms to minuscule script]. *Linguistische Berichte*, 155, 3–21.
- Chen, J.-Y., & Cherng, R.-J. (2013). The proximate unit in Chinese handwritten character production. *Frontiers in Psychology*, 4(517), doi:10.3389/fpsyg.2013.00517
- Coulmas, F. (1996). Typology of writing systems. In H. Günther, & O. Ludwig (Eds.), *Writing and its use* (pp. 1380–1387). Berlin: De Gruyter. doi:10.1515/978 3110147445.2.9.1380
- Daniels, P. T. (1991). Is a structural graphemics possible? *LACUS Forum*, 18, 528–537.
- Daniels, P. T. (2017). Writing systems. In M. Aronoff & J. Rees-Miller (Eds.), *The handbook of linguistics* (2nd ed., pp. 75–94). Oxford: Wiley-Blackwell. doi:10.1002/9781119072 256.ch5
- Daniels, P. T. (2018). *An exploration of writing*. Bristol: Equinox.
- DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu: University of Hawaii Press.
- Dehaene, S. (2009). *Reading in the brain: The new science of how we read*. London: Penguin.
- Dürscheid, C. (2016). *Einführung in die Schriftlinguistik* [Introduction to Grapholinguistics] (5th ed.). Göttingen: Vandenhoeck & Ruprecht.
- Evertz, M. (2016). Graphematischer Fuß und graphematisches Wort [Graphematic foot and graphematic word]. In U. Domahs, & B. Primus (Eds.), *Handbuch Laut, Gebärde, Buchstabe* [Handbook sound, sign, letter] (pp. 377–397). Berlin: De Gruyter. doi:10.1515/9783110295993-021

- Evertz, M. (2018). *Visual prosody: The graphematic foot in English and German*. Berlin: De Gruyter. doi:10.1515/9783110583441
- Fedorova, L. (2013). The development of graphic representation in abugida writing: The akshara's grammar. *Lingua Posnaniensis*, 55(2), 49–66. doi:10.2478/linpo-2013-0013
- Fuhrhop, N. (2008). Das graphematische Wort (im Deutschen): Eine erste Annäherung [The graphematic word (in German): a first approach]. *Zeitschrift für Sprachwissenschaft*, 27(2), 189–228. doi:10.1515/zfs-2008.010
- Fuhrhop, N., & Buchmann, F. (2009). Die Längenhierarchie: Zum Bau der graphematischen Silbe [The length hierarchy: on the structure of the graphematic syllable]. *Linguistische Berichte*, 218, 127–155.
- Fuhrhop, N., & Peters, J. (2013). *Einführung in die Phonologie und Graphematik* [Introduction to phonology and graphematics]. Metzler: Stuttgart.
- Fuhrhop, N., & Schmidt, K. (2014). Die zunehmende Profilierung der Schreibsilbe in der Geschichte des Deutschen [The increasing profiling of the written syllable in the history of German]. *Beiträge zur Geschichte der deutschen Sprache und Literatur*, 136(4), 538–568. doi:10.1515/bgsl-2014-0047
- Gallmann, P. (1985). *Graphische Elemente der geschriebenen Sprache: Grundlagen für eine Reform der Orthographie* [Graphic elements of written language: foundations for an orthography reform]. Berlin: De Gruyter. doi:10.1515/9783111630380
- Garbe, B. (1985). Graphemtheorien und mögliche strukturmodelle zur beschreibung der orthographie [Grapheme theories and possible structural models for the description of orthography]. In G. Augst (Ed.), *Graphematik und Orthographie* [Graphematics and orthography] (pp. 1–21). Frankfurt a. M.: Peter Lang.
- Gelb, I. J. (1969). *A study of writing* (2nd ed., 3rd impression). Chicago: University of Chicago Press.
- Glück, H. (2016). Graphem [Grapheme]. In H. Glück & M. Rödel (Eds.), *Metzler Lexikon Sprache* [Metzler encyclopedia language] (5th ed., pp. 252–253). Stuttgart: Metzler.
- Gnanadesikan, A. E. (2017). Towards a typology of phonemic scripts. *Writing Systems Research*, 9(1), 14–35. doi:10.1080/17586801.2017.1308239
- Günther, H. (1988). *Schriftliche Sprache: Strukturen geschriebener Wörter und ihre Verarbeitung beim Lesen* [Written language: structures of written words and their processing during reading]. Berlin: De Gruyter. doi:10.1515/9783110935851
- Harris, R. (1994). Semiotic aspects of writing. In H. Günther, & O. Ludwig (Eds.), *Writing and its use* (pp. 41–48). Berlin: De Gruyter. doi:10.1515/978311011293.1.1.41
- Harris, R. (2005). Schrift und linguistische Theorie [Writing and linguistic theory]. In G. Grube, W. Kogge, & S. Krämer (Eds.), *Schrift: Kulturtechnik zwischen Auge, Hand und Maschine* [Writing: cultural technique between eye, hand, and machine] (pp. 61–80). München: Wilhelm Fink.
- Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, 86(3), 663–687. doi:10.1353/lan.2010.0021
- Häffner, S. (2009). *Universalien der Schrift und Schriftverarbeitung: zu den schrifttheoretischen Prämissen psychologischer Lesemodelle* [Universals of writing and the processing of writing: on the grapholinguistic premises of psychological models of reading]. Frankfurt a. M.: Peter Lang.
- Heller, K. (1980). Zum Graphembegriff [On the notion grapheme]. In D. Nerius, & J. Scharnhorst (Eds.), *Theoretische Probleme der deutschen Orthographie* [Theoretical problems of German orthography] (pp. 74–108). Berlin: Akademie-Verlag.
- Henderson, L. (1985). On the use of the term 'grapheme'. *Language and Cognitive Processes*, 1(2), 135–148. doi:10.1080/01690968508402075
- Herrick, E. M. (1974). A taxonomy of alphabets and scripts. *Visible Language*, 8(1), 5–32.
- Herrick, E. M. (1994). Of course a structural graphemics is possible! *LACUS Forum*, 21, 413–424.
- Ho, C. S.-H., & Bryant, P. (1997). Learning to read Chinese beyond the logographic phase. *Reading Research Quarterly*, 32(3), 276–289. doi:10.1598/RRQ.32.3.3
- Ho, C. S.-H., Ng, T.-T., & Ng, W.-K. (2003). A 'radical' approach to reading development in Chinese: The role of semantic radicals and phonetic radicals. *Journal of Literacy Research*, 35(3), 849–878. doi:10.1207/s15548430jlr3503_3
- Holenstein, E. (1983). Double articulation in writing. In F. Coulmas, & K. Ehlich (Eds.), *Writing in focus* (pp. 45–62). Berlin: De Gruyter. doi:10.1515/9783110822830.45
- Klinkenberg, J.-M., & Polis, S. (2018). On scripturology. *Signata: Annals of Semiotics*, 9, 57–102.
- Kohrt, M. (1985). *Problemgeschichte des Graphembegriffs und des frühen Phonembegriffs* [History of the notion grapheme and the early notion of phoneme]. Berlin: De Gruyter. doi:10.1515/9783111371207
- Kohrt, M. (1986). The term 'grapheme' in the history and theory of linguistics. In G. Augst (Ed.), *New trends in graphemics and orthography* (pp. 80–96). Berlin: De Gruyter. doi:10.1515/9783110867329.80
- Krämer, S. (2003). ›Schriftbildlichkeit: oder: Über eine (fast) vergessene Dimension der Schrift [›Imagery of writing: or: on a(n almost) forgotten dimension of writing]. In S. Krämer, & H. Bredekamp (Eds.), *Bild, Schrift, Zahl* [Image, writing, number] (pp. 157–176). München: Wilhelm Fink.
- Ladd, D. R. (2014). *Simultaneous structure in phonology*. Oxford: Oxford University Press.
- Lindqvist, C. (2001). *Skandinavische Schriftsysteme im Vergleich* [Scandinavian writing systems in comparison]. Berlin: De Gruyter. doi:10.1515/9783110927085

- Lockwood, D. G. (2001). Phoneme and grapheme: How parallel can they be? *LACUS Forum*, 27, 307–316.
- Meletis, D. (2015). *Graphetik: Form und Materialität von Schrift* [Graphetics: form and materiality of writing]. Glückstadt: Werner Hülsbusch.
- Meletis, D. (2017). Review of *Handbuch Laut, Gebärde, Buchstabe*, ed. by U. Domahs & B. Primus. *Zeitschrift für Rezensionen zur germanistischen Sprachwissenschaft*, 9(1-2), 109–115. doi:10.1515/zrs-2017-0019
- Meletis, D. (2018). What is natural in writing? Prolegomena to a Natural Grapholinguistics. *Written Language and Literacy*, 21(1), 52–88. doi:10.1075/wll.00010.mel
- Meletis, D. (2019). *Naturalness in scripts and writing systems: Outlining a Natural Grapholinguistics* (Unpublished doctoral dissertation). University of Graz.
- Mugdan, J. (1990). On the history of linguistic terminology. In H.-J. Niederehe, & E. F. K. Koerner (Eds.), *History and historiography of linguistics: Proceedings of the 4th International Conference on the history of the language sciences (ICHoLS IV)*, Trier, 24–28 August 1987, Vol. 1: *Antiquity–17th century* (pp. 49–62). Amsterdam: John Benjamins. doi:10.1075/sihols.51.1.09mug
- Myers, J. (2019). *The grammar of Chinese characters: Productive knowledge of formal patterns in an orthographic system*. London: Routledge.
- Neef, M. (2005). *Die Graphematik des Deutschen* [The graphematics of German]. Berlin: De Gruyter. doi:10.1515/9783110914856
- Neef, M. (2015). Writing systems as modular objects: Proposals for theory design in grapholinguistics. *Open Linguistics*, 1, 708–721. doi:10.1515/opli-2015-0026
- Nöth, W. (1990). *Handbook of semiotics*. Indiana, IN: Indiana University Press.
- Patel, P. G. (2010). *The Brāhmī writing system: Cross-fertilizing epigraphy, archaeology and linguistics*. New Delhi: Black & White.
- Primus, B. (2003). Zum Silbenbegriff in der Schrift-, Laut- und Gebärdensprache – Versuch einer mediumübergreifenden Fundierung [On the notion of syllable in written, spoken and sign language – attempt at a cross-medium basis]. *Zeitschrift für Sprachwissenschaft*, 23(1), 3–55. doi:10.1515/zfsw.2003.22.1.3
- Primus, B. (2004). A featural analysis of the modern Roman alphabet. *Written Language and Literacy*, 7(2), 235–274. doi:10.1075/wll.7.2.06pri
- Primus, B. (2006). Buchstabenkomponenten und ihre Grammatik [Components of letters and their grammar]. In U. Bredel, & G. Hartmut (Eds.), *Orthographietheorie und Rechtschreibunterricht* [Orthographic theory and spelling instruction] (pp. 5–43). Berlin: De Gruyter. doi:10.1515/9783110921199.5
- Primus, B. (2010). Strukturelle Grundlagen des deutschen Schriftsystems [Structural basics of the German writing system]. In U. Bredel, A. Müller, & G. Hinney (Eds.), *Schriftsystem und Schriffterwerb: Linguistisch – didaktisch – empirisch* [Writing system and literacy acquisition: linguistically, didactically, empirically] (pp. 9–46). Berlin: De Gruyter. doi:10.1515/9783110232257.9
- Qiu, X. (2000). *Chinese writing*. (G. L. Mattos, & J. Norman, Trans.). Berkeley, CA: The Society for the Study of Early China.
- Rezec, O. (2009). *Zur Struktur des deutschen Schriftsystems* [On the structure of the German writing system] (Doctoral dissertation). University of Munich. Retrieved from <https://edoc.ub.uni-muenchen.de/10730/>
- Rezec, O. (2013). Ein differenzierteres Strukturmodell des deutschen Schriftsystems [A more differentiated structural model of the German writing system]. *Linguistische Berichte*, 234, 227–254.
- Rizza, A. (2018, November). *Recurring problems in the application of the term/concept grapheme: reflecting again*. Paper presented at the LSS München/Graphematik-Tagung des Akademieprojekts RuneS “Wege zur Konfiguration der Zeichen-Phonem-Beziehung”, Munich, Germany.
- Rogers, H. (2005). *Writing systems: A linguistic approach*. Oxford: Wiley-Blackwell.
- Saiegh-Haddad, E. & Henkin-Roitfarb, R. (2014). The structure of Arabic language and orthography. In E. Saiegh-Haddad, & R. Malatesha Joshi (Eds.), *Handbook of Arabic literacy: Insights and perspectives* (pp. 3–28). Dordrecht: Springer. doi:10.1007/978-94-017-8545-7_1
- Salomon, R. (2007). Typological observations on the Indic scripts and their relationship with other alphasyllabaries [sic]. In P. G. Patel, P. Pandey, & D. Rajgor (Eds.), *The Indic scripts: Palaeographic and linguistic perspectives* (pp. 25–43). New Delhi: D. K. Printworld.
- Sampson, G. (1994). Chinese script and the diversity of writing systems. *Linguistics*, 32(1), 117–132. doi:10.1515/ling.1994.32.1.117
- Sampson, G. (2015). *Writing systems: A linguistic introduction* (2nd ed). Bristol: Equinox.
- Schmidt, K. (2016). Der graphematische Satz: Vom Schreibsatz zur allgemeinen Satzvorstellung [The graphematic sentence: from the sentence in writing to a general notion of sentence]. *Zeitschrift für germanistische Linguistik*, 44(2), 215–256. doi:10.1515/zgl-2016-0011
- Schmidt, K. (2018). *Phonographie und Morphographie im Deutschen. Grundzüge einer wortbasierten Graphematik* [Phonography and Morphography in German: basics of a word-based graphematics]. Tübingen: Stauffenburg.
- Share, D. L., & Daniels, P. T. (2016). Aksharas, alphasyllabaries, abugidas, alphabets and orthographic depth: Reflections on Rimzhim, Katz, and Fowler (2014). *Writing Systems Research*, 8(1), 17–31. doi:10.1080/17586801.2015.1016395
- Sproat, R. (2000). *A computational theory of writing systems*. Cambridge, UK: Cambridge University Press.

- Taha, H. Y. (2013). Reading and spelling in Arabic: Linguistic and orthographic complexity. *Theory and Practice in Language Studies*, 3(5), 721–727. doi:10.4304/tpls.3.5.721-727
- Valkama, M. (2017, August). *How to segment writing into comparable units?* Paper presented at the 11th International AWLL Workshop, Nagoya, Japan. Retrieved from https://www.academia.edu/34468726/How_to_segment_writing_into_comparable_units_11th_AWLL_workshop_Nagoya_29.8.2017_.
- Waller, R. H. W. (1991). Typography and discourse. In R. Barr, M. L. Kamil, P. D. Pearson & P. Rosenthal (Eds.), *Handbook of reading research*, (Vol. 2, pp. 341–380). New York: Longman.
- Weingarten, R. (2011). Comparative graphematics. *Written Language and Literacy*, 14(1), 12–38. doi:10.1075/wll.14.1.02wei
- Winkel, H., & Iemwanthong, K. (2010). Reading and spelling acquisition in Thai children. *Reading and Writing*, 23(9), 1021–1053. doi:10.1007/s11145-009-9194-6