What is a *grapheme*? Do we need it?

Re-evaluating one of grapholinguistics' core notions

Dimitrios Meletis
University of Graz

WRITING SYSTEMS: PAST, PRESENT (… AND FUTURE?)
AWLL's 11th International Workshop on Writing Systems
Nanzan University, Nagoya, Japan
August 29th, 2017

UNI GRAZ

Outline

1. The grapheme in grapholinguistics: *Three situations*
    a. Opponents: *There is no grapheme*
    b. Non-committers/neutrals: *Using, but not defining* or *Not using it at all*
    c. Proponents: *There is a grapheme, but what is it? – Referential vs. analogical*
2. The grapheme as a relation: *Basic shapes* and *linguistic units*
3. A special challenge: *What is allography, then?*
4. Preliminary conclusion: *The grapheme as a cross-grapholinguistic basic relation*
5. References

1. The *grapheme* in grapholinguistics

a. Opponents: *There is no grapheme*

- **claim:** positing a concept/unit 'grapheme' analogous to other emic units is impossible because writing differs fundamentally from language (in that it is conscious, while language is not, cf. DANIELS 1991); also, the term has become so overladen with meaning, it is useless:
- "'Grapheme' […] has had so many different interpretations that in writing systems theory it is meaningless" (SHARE/DANIELS 2016: 23)

b. Non-committers/neutrals: *Using & not defining, not using at all*

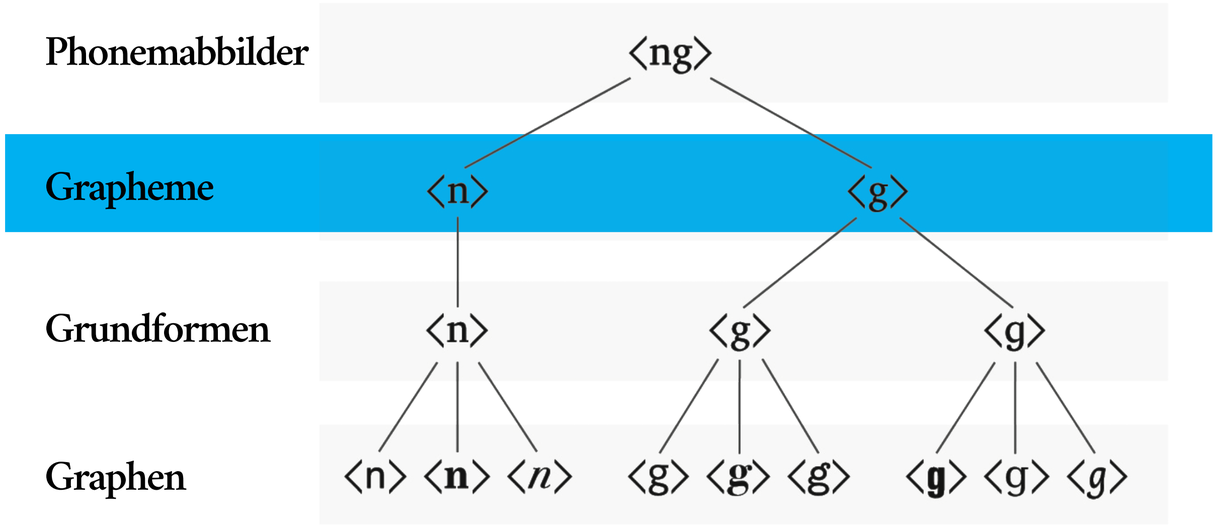
- some use 'grapheme' (*without* defining it) in very different meanings, often in the vague sense of 'the smallest unit of a writing system' (functional, material?); this happens in (often not theoretically-oriented, sometimes non-linguistic) works on many writing systems (Chinese, cf. CHEN & CHERNG 2013; Thai, cf. WINSKEL & IEMWANTHONG 2010; Arabic, cf. TAHA 2013)
- some acknowledge the (possible) existence and value of a 'grapheme' but choose not to make it a crucial unit it in their analyses, working with 'letter' instead (NEEF 2005; PRIMUS 2006)

1. The *grapheme* in grapholinguistics

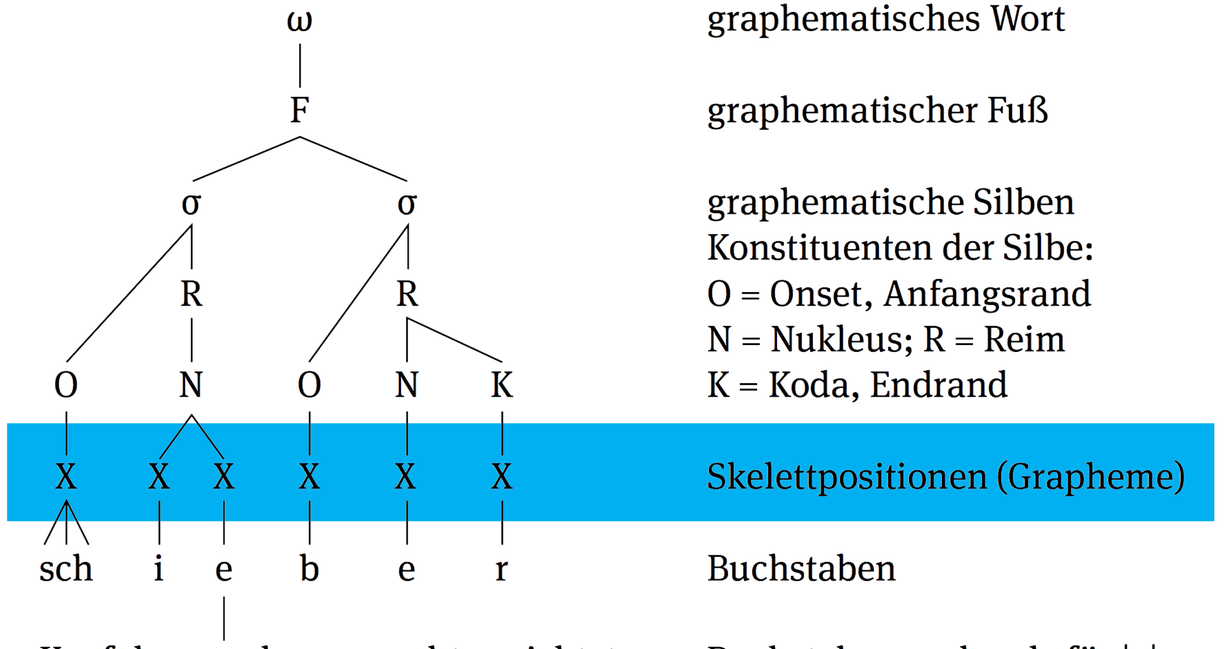
c. Proponents: *There is a grapheme, but what is it?*

- two different definitions of the grapheme that depend on what relation between speech and writing is assumed: **dependency hypothesis** (= referential) vs. **autonomy hypothesis** (= analogical)
- **Referential:** the grapheme corresponds to a phoneme; thus, in German, for example, <sch> is one grapheme because it corresponds to one phoneme, /ʃ/; <v>, <ph> and <f> are allographs of the grapheme for/f/ – but writing is not just a depiction of speech
- **Analogical:** the method of compiling the grapheme inventory of a writing system is the same as the method of establishing the phonemes of a language, because a writing system is its *own* system and graphotactic regularities often don't follow phonological regularities (cf. GÜNTHER 1988: 77); thus, minimal pairs serve as evidence for graphemes; however, some challenges arise:
  - technically, because of a pair like <denkt> and <deckt> (cf. REZEC 2013: 231), both <n> and <c> are graphemes of German; however, even in the analogical view, the status of <c> as a grapheme is contested as its distribution is very limited and it does not occur on its own (without <k>, <h> etc.) in any *native* graphematic words of German (cf. FUHRHOP/PETERS 2013: 204)
  - is <sch> a grapheme in this view, too? – there *are* minimal pairs: <Schaum> vs. <Baum>

1. The *grapheme* in grapholinguistics



(cf. REZEC 2009: 99)



(cf. BERG, PRIMUS & WAGER 2016: 351)

Please note that this discussion as well as the ensuing ideas and terms are largely **Eurocentric,** more precisely **alphabetocentric.**

So how do these views and their concepts hold up if we look at a *diverse range* of writing systems, attempting to find *universal traits*?

2. The '*grapheme*' as a relation

- the grapheme *is* an emic unit; it always **refers** to language – however, it *is* not language itself, it is only language by extension, so it is logically dependent on linguistic information (*not* only linguistic units)
- if differences in writing relate with differences on a linguistic level, it is because the units of writing systematically relate to the units of language
  - however, if units don't differ on *any* linguistic level and still differ in writing – is the difference graphematic?, e.g. <Typographie> vs. <Typografie> (= **orthographic**), |Allograph| vs. |*Allograph*| (= **graphetic**)
- sometimes units of writing differentiate meaning without referring to a linguistic unit on their own
- also, sometimes units of writing (such as punctuation marks) refer to linguistic functions rather than units
- as there is type mixing in almost all writing systems, the systems make use of different types of linguistic information and thus, various types of graphemes are at work within given writing systems, not only graphemes that refer to the basic unit of operation of a writing system (alphabet = phoneme, morphographic writing system = morpheme, etc.); this does *not* mean that there is no such thing as a grapheme

2. The '*grapheme*' as a relation

grapheme (= graphematic relation)

basic shape(s)

= abstract visual units, 'skeletons'

- constrained by human eyes, brains, hands, and available writing instruments and material (WATT 1999)
- there are universal regularities and systematic traits at this level (CHANGIZI & SHIMOJO 2005; CHANGIZI ET AL. 2006)

distinguishing meaning

linguistic unit(s)

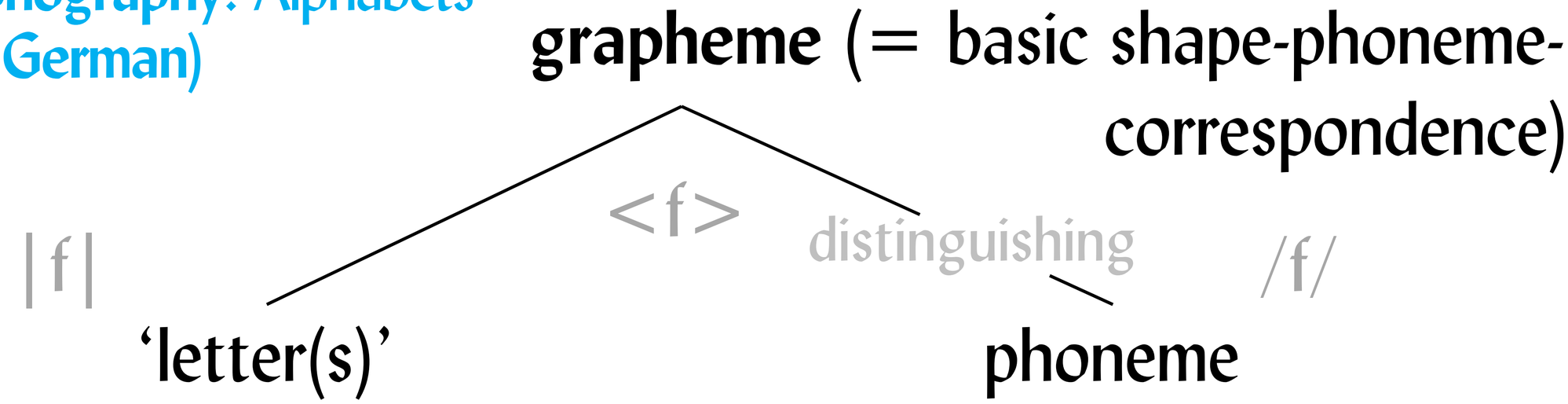
- of linguistic levels that share the feature that they are (relatively) closed inventories

linguistic function(s)

- other functions: differentiating meaning, indicating syntactic, prosodic etc. information

2. The '*grapheme*' as a relation

a. Phonography: Alphabets (cf. German)



- Is 'letter' a graphetic or a graphematic term? (For BERG, PRIMUS & WAGNER 2016, it is graphematic.) Is it alphabet-specific?
- If we take the analogical view to be primary (for German, at least), then why do some units like |c| that do not correspond with a phoneme have an unusual distribution and only occur together with other units? (And in combination, they then do refer to a phoneme.)
- If we take both views to be partially correct, how do we reconcile them? If we only take units that distinguish meaning as graphemes, how does that take into account that they also refer to phonemes?

2. The '*grapheme*' as a relation

b. Phonography: Abjads (cf. Arabic)

grapheme (= basic shape-phoneme-correspondence)

'letter(s)' (?) — consonant(s) (+ some vowels)

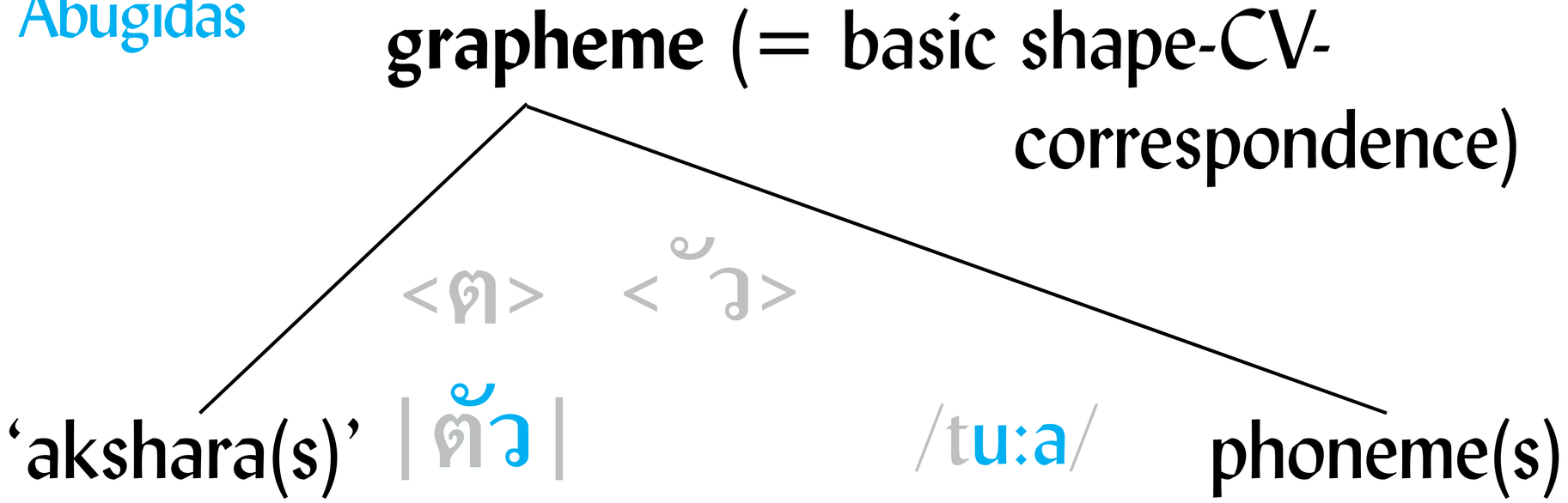
|ب| — /b/

- How can it be explained in this model that some vowel phonemes are indicated and some are not? Does it even have to be explained?
- Do the optional vowel signs have graphematic status or not? They differentiate meaning and refer to phonemes.

2. The '*grapheme*' as a relation

c. Phonography: Abugidas (cf. Thai)

Could PRIMUS' et al. suprasegmental model be valuable here, too?

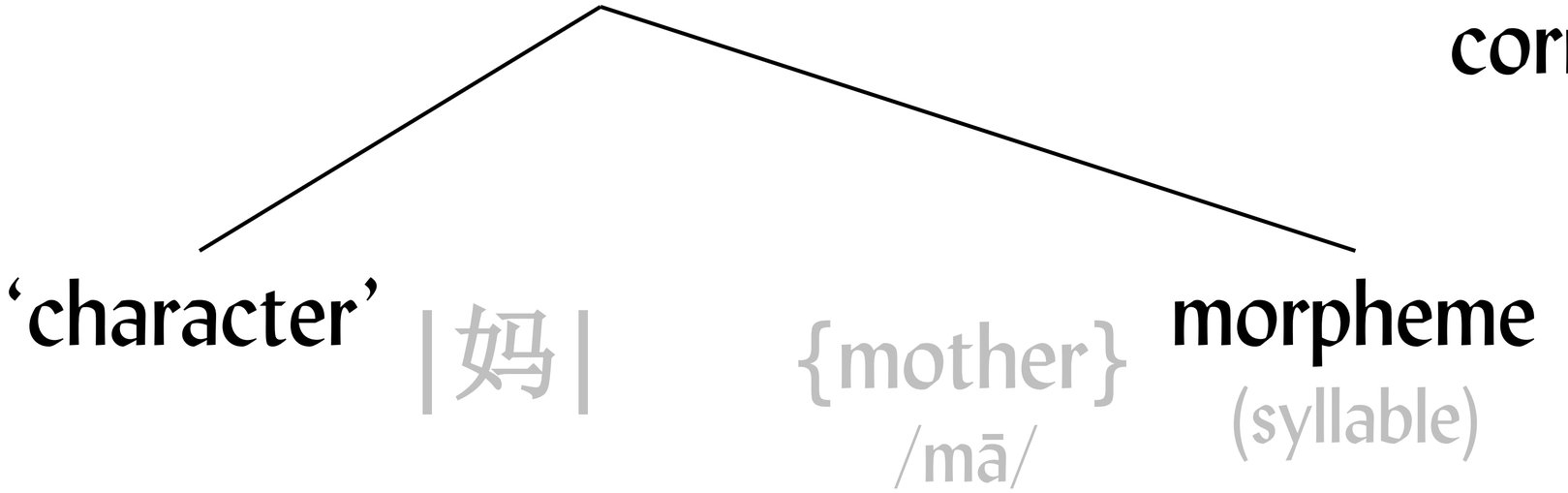


– With the 'unmarked' aksharas referring to C_V sequences and the vowel 'diacritics' being obligatory, what is the status of these units? Are the unmarked aksharas graphemes and the aksharas with diacritics derived graphemes? Is there one $C_{(V)}$ grapheme and a few consistent V graphemes? Is economy an important factor in positing a grapheme inventory for an abugida system (or any system, cf. analogical graphemes in German that combine to referential graphemes)? Do bound vs. free graphemes (cf. ROGERS 2005) have a different status?

2. The '*grapheme*' as a relation

d. Morphography (cf. Chinese)

grapheme (= basic shape-morpheme-syllable-correspondence)



Higher levels presuppose lower levels: morphemes have a phonemic representation.

- Again: Is 'character' a graphetic term? What inventories should it be used for?
- If graphemes are not used for their morphemic information, but instead for their phonological representation, is this still the same type of grapheme?
- Are the characters themselves the smallest graphematic units or are the phonetic and semantic components graphemes (cf. DeFrancis 1989)?

3. A special challenge: *Allography*

In grapholinguistics there are two (maybe three) types of *allography*:

Graphetic allographs

1. **different graphs** that are associated with the same basic shape: <g>, <g>, <g>, <g>
2. **different basic shapes** that are associated with the same grapheme: <a>, <ɑ>, <g>, <*g*>
   - ones that are in complementary distribution, e.g. positional variants in Greek or Arabic: <σ>, <ς>, <ـبـ>, <ـب>, <ب>
   - free choices that are "dialectal"; they are associated with an inventory: one font uses <a>, the other <ɑ>
   - allographs across writing systems due to diverging developments or political reasons, cf. <戶> vs. <户> vs. <戸>
   - for the ones that can be exchanged: exchanging them does not lead to a difference in meaning and also not to an error, a graphematically unlicenced form

Graphematic allographs

- graphematic allographs (or *allographemes*?) are variants of one grapheme where there are either minimal pairs without a difference in meaning: <Typographie> vs. <Typografie>
- or they represent the possibilities of how a phoneme can be encoded in writing: /f/ can be written <v>, <f>, <ph>; there are *no* minimal pairs with <v> and <f> referring to /f/ (= referential), but there is <Phase> vs. <Vase>
- they occur syntagmatically with different referents
- = how can graphematic allographs be allographs and separate graphemes at the same time? How often does this occur in the writing systems of the world?

4. Conclusion

Previously on "The grapheme discussion"

- **both the (1) analogical as well as the (2) referential view are not adequate (on their own):**
  - (1) graphemes can't be assumed analogously to phonemes, because unlike phonemes, they refer to something – they are "signs of signs"; also, this view would exclude graphemes that refer not to spoken language;
  - (2) graphemes do not only represent phonemes; sometimes, they might just differentiate meaning with a unit that does not neatly correspond to any linguistic unit; they might also refer to other linguistic levels than the phonological
- any concept of grapheme needs to account for *all* writing systems, not just phonographic ones

4. Conclusion

at least **two criteria** need to be considered when asking if a a written unit is a grapheme or not:

1. it needs to have a linguistic value by fulfilling one of these conditions (in this order):
   a. first, by differentiating meaning as in <deckt> vs. <denkt>; here, <c> does not relate to a phoneme on its own but it is functional in that it differentiates meaning – again, how often does this occur in writing systems?
   b. by being a relation between a basic shape (or basic shapes) and a linguistic unit (phoneme(s), syllable, morpheme); if two basic shapes refer to the same linguistic unit (= no different meaning), we speak of allography
   c. by referring to a linguistic function or linguistic information of some other kind; in the case of punctuation syntactical or prosodic information, for example
2. it needs to be the smallest possible "unit" of writing; this means that if <a> and <h> are already graphemes, we do not need a grapheme <ah> – the number of polygraphs should be kept to a minimum and graphemes combine to larger graphematic units (**economy**)

Graphemes for which not all of these criteria are met, are **exceptions**. The definition for the grapheme should not be based on exceptions.

5. References

Berg, Kristian & Beatrice Primus & Lutz Wagner (2016): Buchstabenmerkmal, Buchstabe, Graphem. In: Ulrike Domahs & Beatrice Primus (eds.): *Handbuch Laut, Gebärde, Buchstabe* (Handbücher Sprachwissen 2.) Berlin, Boston: De Gruyter, 337-355.

Changizi, Mark A. & Shinsuke Shimojo (2005): Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society B* 272: 267-275.

Changizi, Mark A., Qiong Zhang, Hao Ye & Shinsuke Shimojo (2006): The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *The American Naturalist* 167.5: E117-E139.

Chen, Jenn-Yeu & Rong-Ju Cherng (2013): The proximate unit in Chinese handwritten character production. *Frontiers in Psychology* 4. DOI: 10.3389/fpsyg.2013.00517.

Daniels, Peter T. (1991): Is a structural graphemics possible? *LACUS Forum* 18: 528-537.

DeFrancis, John (1989): *Visible speech: the diverse oneness of writing systems*. Honolulu: University of Hawaii Press.

Fuhrhop, Nanna & Jörg Peters (2013): *Einführung in die Phonologie und Graphematik*. Stuttgart: Metzler.

Günther, Hartmut (1988): *Schriftliche Sprache: Strukturen geschriebener Wörter und ihre Verarbeitung beim Lesen.* (*Konzepte der Sprach- und Literaturwissenschaft*, 40.) Tübingen: Niemeyer.

Neef, Martin (2005): *Die Graphematik des Deutschen.* (*Linguistische Arbeiten*, 500.) Tübingen: Niemeyer.

Primus, Beatrice (2006): Buchstabenkomponenten und ihre Grammatik. In: Bredel, Ursula & Günther Hartmut (Hrsg): *Orthographietheorie und Rechtschreibunterricht*, 5-43. Tübingen: Niemeyer.

Rezec, Oliver (2009): *Zur Struktur des deutschen Schriftsystems.* Ludwig-Maximilians-Universität München Dissertation.

Rezec, Oliver. 2013. Ein differenzierteres Strukturmodell des deutschen Schriftsystems. *Linguistische Berichte* 234: 227-254.

Rogers, Henry (2005): *Writing systems: a linguistic approach*. Malden: Blackwell.

Share, David L. & Peter T. Daniels (2016): Aksharas, alphasyllabaries, abugidas, alphabets and orthographic depth: Reflections on Rimzhim, Katz and Fowler (2014). *Writing Systems Research* 8.1: 17-31.

Taha, Haitham Y. (2013): Reading and Spelling in Arabic: Linguistic and Orthographic Complexity. *Theory and Practice in Language Studies* 3.5: 721-727.

Watt, W. C. (1999): How to recognize extraterrestrial symbols, when and if. *Semiotica* 125.1-3: 75-82.

Winskel, Heather & Kanyarat Iemwanthong (2010): Reading and spelling acquisition in Thai children. *Reading and Writing* 23.9: 1021-1053.

Thank You!