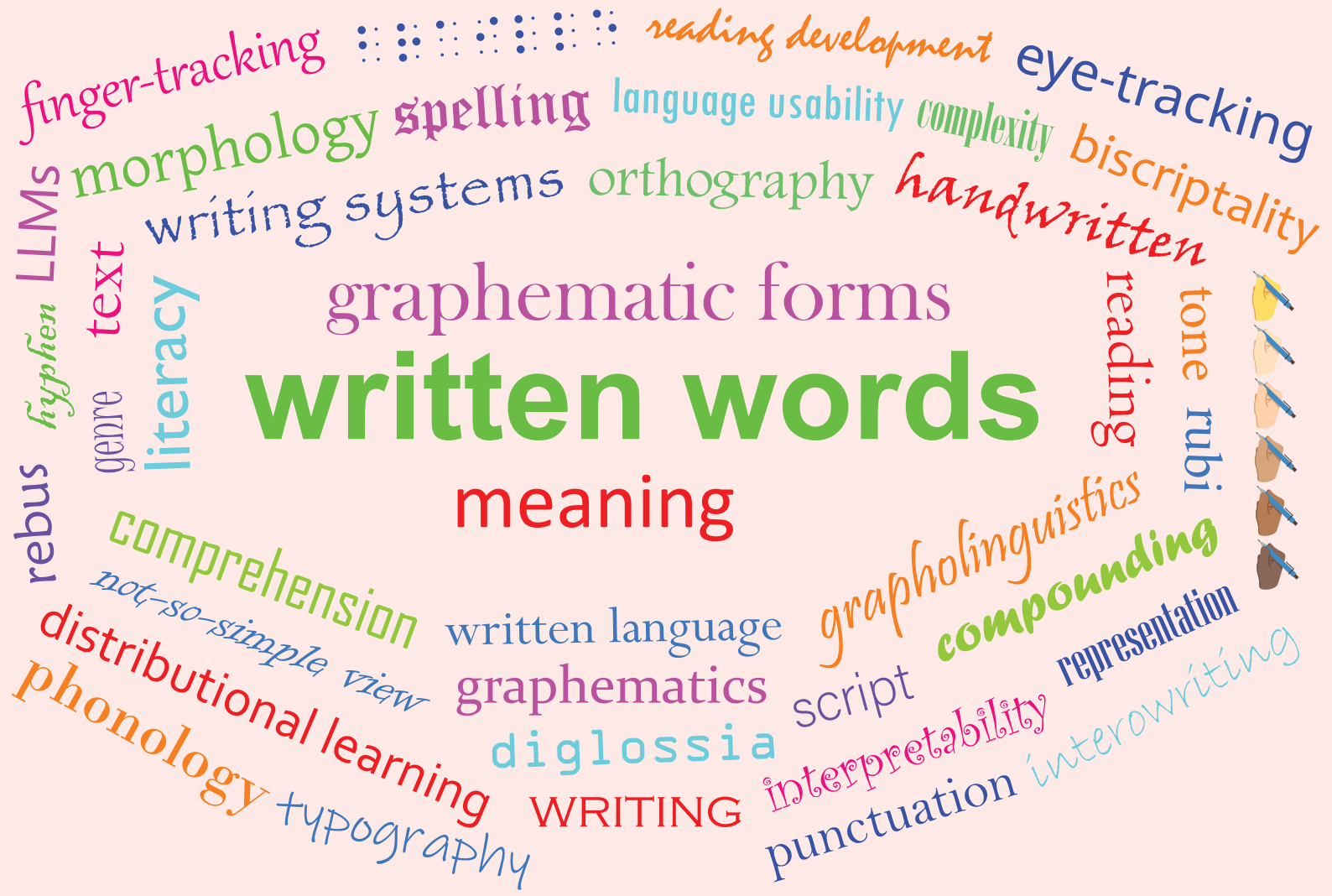
Written words:
From graphematic forms to meaning



15th International Workshop on Writing Systems and Literacy

29-31 October 2025

Italian National Research Council (CNR), Institute for Computational Linguistics "Antonio Zampolli", Pisa, Italy

Association for Written Language and Literacy
Ассоциация письменного языка и письменности
Ενωση για τη γραπτη γλωσσα και τη βασικη εκπαιδευση
书面语言和识字协会
Association for Written Language and Literacy

Istituto di Linguistica Computazionale
"Antonio Zampolli"
Consiglio Nazionale delle Ricerche

Written words: From graphematic forms to meaning

Association for Written Language and Literacy's 15th International Workshop on Writing Systems and Literacy (AWLL15)

29-31 October 2025

Italian National Research Council (CNR), Institute for Computational Linguistics "Antonio Zampolli" Pisa, Italy

Local organizer

Vito Pirrelli (CNR Pisa, Italy)

Programme committee

Terry Joyce (Tama University; Temple University Japan)
Lynne Cahill (University of Sussex, UK)
Dorit Ravid (Tel Aviv University, Israel)
Vito Pirrelli (CNR Pisa, Italy)

AWLL15 gratefully acknowledges a donation from Amalia E. Gnanadesikan.

Day 1: Wednesday, 29 October 2025

08:30 – 09:00	Check-in
09:00 - 09:30	**Opening remarks**
09:30 – 10:30	**Oral session 1**
10:30 – 11:00	Break
11:00 – 12:00	**Oral session 2**
12:00 – 13:00	Lunch
13:00 – 14:30	**Oral session 3**
14:30 – 15:00	Break
15:00 – 16:30	**Oral session 4**
16:30 – 17:00	Break
17:00 – 18:30	**Symposium: [title?]**
18:30 – 19:00	Transit gap
19:00 – 22:00	**Walking tour to reception venue**

Day 2: Thursday, 30 October 2025

09:00 – 10:30	**Oral session 5**
10:30 – 11:00	Break
11:00 – 12:00	**Keynote 1**
12:00 – 13:00	Lunch
13:00 – 14:15	**Poster session 1**
14:15 – 14:45	Break
14:45 – 16:15	**Oral session 6**
16:15 – 16:45	Break
16:45 – 17:45	**Oral session 7**
17:45 – 18:00	Break
18:00 – 18:45	**AWLL business meeting**
18:45 – 20:00	Transit gap
20:00 – 22:00	**Conference dinner**

Day 3: Friday, 31 October 2025

09:00 – 10:30	**Oral session 8**
10:30 – 11:00	Break
11:00 – 12:00	**Keynote 2**
12:00 – 13:00	Lunch
13:00 – 14:15	**Poster session 2**
14:15 – 14:45	Break
14:45 – 16:15	**Oral session 9**
16:15 – 16:45	Break
16:45 – 17:45	**Oral session 10**
17:45 – 18:00	Break
18:00 – 18:45	**Panel discussion**
18:45 – 19:00	**Closing remarks** [plus group photograph]

Day 1: Wednesday, 29 October 2025

08:30 – 09:00	Check-in
09:00 - 09:30	**Opening remarks**
09:30 – 10:30	**Oral session 1**
09:30 – 10:00	Lindsay Harris Does the complexity of Unified English Braille interfere with self-teaching?
10:00 – 10:30	Keisuke Honda Graphematic representation in Japanese braille and print: Implications for grapholinguistics
10:30 – 11:00	Break
11:00 – 12:00	**Oral session 2**
11:00 – 11:30	Lynne Cahill How many ways can you spell that? Intra-document variation in Middle English documents
11:30 – 12:00	Sara Budts, Yoshi Malaise, & Rik Vosters Standardization and orthographic variation in Late Modern Dutch witness depositions
12:00 – 13:00	Lunch
13:00 – 14:30	**Oral session 3**
13:00 – 13:30	Rana Yassin & Yasmin Shalhoub-Awwad Characteristics of early spelling errors in Arabic: The impact of visual-orthographic features
13:30 – 14:00	Nancy Joubran-Awadie & Yasmin Shalhoub-Awwad How does morphological awareness before and after the onset of formal reading instruction affect later reading outcomes in a rich morphological language?
14:00 – 14:30	Deia Ganayim Writing units or decades first in two digit numbers dictation tasks: The case of Arabic—an inverted writing system
14:30 – 15:00	Break
15:00 – 16:30	**Oral session 4**
15:00 – 15:30	Marta Guidotti The visual and typographic dimension of language: Analysis and clustering of languages, based on language design features
15:30 – 16:00	Dimitrios Meletis “The worst writing system in history”: Public attitudes toward the learnability of writing systems
16:00 – 16:30	Terry Joyce, Hisashi Masuda, & Chikako Fujita Japanese *rubi*: Layered representation of meaning
16:30 – 17:00	Break

17:00 – 18:30	**Symposium: Perspectives on text reading and comprehension**
17:00 – 17:30	Stefano Rastelli Plain language: A psycholinguistic approach (no forest for the trees yet)
17:30 – 18:00	CNR-ILC ItaliaNLP Lab Profiling linguistic complexity: From human perception to language model interpretability
18:00 – 18:30	CNR-ILC Comphys Lab Reading development is about multimodal coordination and linguistic awareness: An integrative, not-so-simple view of reading
18:30 – 19:00	Transit gap
19:00 – 22:00	**Walking tour to reception venue**

Day 2: Thursday, 30 October 2025

09:00 – 10:30	**Oral session 5**
09:00 – 09:30	Naama Evanhaim, Daphna Lavi-Mudrik & Amalia Bar-On Rethinking morphology-based reading in Hebrew: New findings using a finger-tracking paradigm
09:30 – 10:00	Galit Ben-Zvi & Amalia Bar-On Unpacking the role of genre and word-level information in reading Hebrew
10:00 – 10:30	David L. Share & Elinor Saiegh-Haddad Arabic and Hebrew: Fraternal but not identical twins
10:30 – 11:00	Break
11:00 – 12:00	**Keynote 1**
	Elinor Saiegh-Haddad The triangulation of linguistic, orthographic, and diglossic factors in reading development in Arabic
12:00 – 13:00	Lunch
13:00 – 14:15	**Poster session 1**
	Aleksandra Twardokęs Visual languages of branding systems. The case of branding system of the Zooba restaurant Rawan Abdulmonem M Almuzaini What can biscriptality offer in the realm of social media? Noboru Yoshioka (How) do underdeveloped writing systems distort grammar? Witten Burushaski ‘words’ in transcription
14:15 – 14:45	Break

14:45 – 16:15	**Oral session 6**
14:45 – 15:15	Svetlana Alexeeva & Alisa Lezina Processing letter position in transparent scripts: Insights from Korean hangul
15:15 – 15:45	Gordon Berthin Comparing the syllabary and logography models of the Rongorongo glyphic script of Easter Island (Rapa Nui)
15:45 – 16:15	Li (Mark) Bai Addressing complexity in learning to read the Manchu writing system
16:15 – 16:45	Break
16:45 – 17:45	**Oral session 7**
16:45 – 17:15	Jonas Romstadt Breaking written words
17:15 – 17:45	Jordi Fortuny Ambiguity and the creation and evolution of writing systems
17:45 – 18:00	Break
18:00 – 18:45	**AWLL business meeting**
18:45 – 20:00	Transit gap
20:00 – 22:00	**Conference dinner**

Day 3: Friday, 31 October 2025

09:00 – 10:30	**Oral session 8**
09:00 – 09:30	Anna Melnikova & John David Storment Emojis and ✍️ing: What counts?
09:30 – 10:00	David Osgarby Pictish symbols: Letting graphetics inform graphematic
10:00 – 10:30	Avital Braun, Dorit Ravid & Elitzur Dattner Compounding patterns in Hebrew writing: A developmental study
10:30 – 11:00	Break
11:00 – 12:00	**Keynote 2**
	Alessandro Lenci How forms shape meaning: What distributional learning can and can't do in the era of LLMs
12:00 – 13:00	Lunch
13:00 – 14:15	**Poster session 2**
	Niamh Kelly & Keisuke Honda The Japanese language learner in the digital age: Re-evaluating the need for handwriting kanji

	Daniel Zagar & Teng Guo The internal structure of writing systems: A cognitive approach Alice Mazzilli Interowriting Niklas Reinken Visible grammatical structures in handwritten texts in German Kathleen Carroll The relationship of writing beliefs to quality and understanding in STEM and non-STEM writing
14:15 – 14:45	Break
14:45 – 16:15	**Oral session 9**
14:45 – 15:15	Sonali Nag Marking the differences: What precursor skills to reading comprehension tell us
15:15 – 15:45	Georgia Niolaki & Aris Terzopoulos Introduction to the Spelling Profile Assessment (SPA): A new spelling test for English primary school children
15:45 – 16:15	Wieke Harmsen, Roeland van Hout, Catia Cucchiarini, & Helmer Strik Towards automated literacy diagnostics: Insights from reading and spelling error analysis
16:15 – 16:45	Break
16:45 – 17:45	**Oral session 10**
16:45 – 17:15	Rosso Manuel Senesi A dynamical framework for graphematic feature emergence: Evidence from error patterns in children's handwriting
17:15 – 17:45	
17:45 – 18:00	Break
18:00 – 18:45	**Panel discussion**
18:45 – 19:00	**Closing remarks** [plus group photograph]

Does the complexity of Unified English Braille interfere with self-teaching?

Lindsay Harris | Northern Illinois University, USA

Purpose. Limited data suggest rates of dyslexia are higher in blind children who read Unified English Braille than in sighted children who read printed English. In the present study, we investigate whether the complex braille code might fuel this disparity by reducing instances of successful decoding and self-teaching during reading development. If blind braille readers underperform sighted print readers on a measure of phonological decoding, it could signal the braille code interferes with the phonological decoding self-teaching loop and increases rates of dyslexia diagnosis in the braille-reading population.

Method. Twenty-two adult print readers and 29 adult braille readers participated in the study. Phonological decoding was assessed with an orthographically based pseudohomophone recognition test that requires participants to silently read letter strings and judge whether, if read aloud, they would sound like real English words (e.g., *kake*) or not (e.g., *threp*). There were 24 items on the assessment, which took approximately 10 minutes to complete.

Results. We conducted a multiple linear regression controlling for age (which significantly differed between groups) of the effect of group (braille, print) on phonological decoding. Results indicated that braille readers performed significantly better than print readers on the phonological decoding assessment, $F(1, 49) = 10.28$, $p = 0.002$.

Conclusions. The finding of superior phonological decoding skills in braille readers runs contrary to our hypothesis. The results suggest that factors other than script, such as a lack of exposure to environmental print and high rates of preterm birth in children born blind, contribute to the prevalence of dyslexia in blind children.

Graphematic representation in Japanese braille and print: Implications for grapholinguistics

Keisuke Honda | Dublin City University, Ireland

Braille is a system of tactile signs formed by raised dots. It is used by blind and visually impaired individuals worldwide to read and write in numerous languages. While braille can be used for sign-by-sign transliteration of print texts (e.g., English Grade 1), the actual braille representation of a language often contains structural features absent from the print representation of the same language (e.g., English Grade 2). Hence, braille and print constitute related but distinct writing systems. Nevertheless, braille remains largely underexplored in grapholinguistics, despite growing recognition that it deserves focused research (Englebretson et al. 2023; Iyengar 2024).

This paper aims to compare the graphematic representations of Japanese braille and print to demonstrate the significance of such cross-modal comparisons for the linguistic analysis of writing systems. Drawing on the conventions of the Braille Authority of Japan (2018), it will highlight two key structural features of Japanese braille representation. First, the braille signs are segmentally encoded and moraically organised (1a), differing from the moraic *kana* signs (1b) and the morphographic *kanji* signs (1c) employed in print. Second, the braille words use blank spaces to separate lexemes as well as trimoraic or longer compound constituents (2a), unlike the unspaced representation in kana (2b) and kanji (2c).

(1) Example: /teN/

a. ⠿⠄⠄ ([を /t/ + え /e/] + ん /N/)

b. てん (て /te/ + ん /N/)

c. 点 (点 {teN} ‘point’)

(2) Example: NIHON+TENJI ‘Japanese braille’

a. ⠇⠄⠄⠀⠿⠄⠔⠞ (⠇⠄⠄ NIHON + ⠀ <BLANK> + ⠿⠄⠔⠞ TENJI)

b. にほんてんじ (にほん NIHON + てんじ TENJI)

c. 日本点字 (日本 NIHON + 点字 TENJI)

The comparison suggests that Japanese braille is a lesser-known type of phonographic writing system. The paper will elaborate on this point and explore the benefits and challenges of incorporating such analysis into grapholinguistics.

Braille Authority of Japan. (2018). *Nihon tenji hyōkihō 2018-nen ban* [日本点字表記法 2018 年版 ‘Japanese braille notation, 2018 edition’]. Hakubunkan Shinsha.

Englebretson, R., Holbrook, M. C., & Fischer-Baum, S. (2023). A position paper on researching braille in the cognitive sciences: Decentering the sighted norm. *Applied Psycholinguistics*, *44*, 400–415. https://doi.org/10.1017/S0142716423000061

Iyengar, A. (2024). *Braille as writing: Evidence and lessons for grapholinguistic theory* [Conference presentation]. Grapholinguistics in the 21st Century, 2024, Università Ca’ Foscari, Venice, Italy. https://youtu.be/LkwfCXVfB78?si=-zJl6ZDdeoIbculp

How many ways can you spell that?
Intra-document variation in Middle English documents

Lynne Cahill | University of Sussex, UK

The standardisation of English spelling happened over a long period, probably beginning some time during the thirteenth century. An accepted set of spelling conventions had more or less been established by the middle of the eighteenth century, when Johnson published his dictionary, using mostly the forms we still use today. Exactly how the process progressed and spread, however, is still the matter of some debate.

Cahill (2023) described an analysis of the MELD corpus of local administrative documents dating between 1399 and 1525 (Stenroos et al 2017). This analysis showed regional differences in the forms used but also, surprisingly, showed that the acceptance that there should be a single spelling for each word did not appear to have been established by the latest documents, which date from the first quarter of the sixteenth century. This is indicated by continued intra-document variation in spellings.

In this paper, we make a comparison between the documentary texts in the MELD corpus and the wider range of genres of texts available in the MEG-C corpus (Middle English Grammar). Following Stenroos's (findings that certain developments in English spelling (e.g. the loss of characters such as thorn) appear to have progressed faster in documentary texts than in literary texts, we hypothesise that the non-documentary texts in MEG-C will, similarly, show a slower move towards eliminating intra-document variant spellings for the same word.

Provisional findings indeed support this hypothesis, with 58% of texts showing only a single variant of the words studied, compared to 73% in the documentary texts. This should be viewed with caution, due to a slightly earlier time period being involved, so further comparisons, taking into account exact timings, locations as well as additional variables will be presented, giving the clearest picture to date of this particular feature of spelling standardisation.

"Middle English Grammar" Manuscripts Online (www.dhi.ac.uk, version 1.0, 20 June 2025), https://www.dhi.ac.uk/manuscriptsonline/resources/mg/

Stenroos, Merja (2020) Regional variation and supralocalization in late medieval English, in Stenroos and Thengs (eds. 2020), pp. 95-128.

Stenroos, Merja, Kjetil V. Thengs and Geir Bergstrøm (2017), *A Corpus of Middle English Local Documents (MELD), version 2017.1*. University of Stavanger. http://www.uis.no/meld (Accessed: 31 December 2020).

Standardization and orthographic variation in Late Modern Dutch witness depositions

Sara Budts, Yoshi Malaise & Rik Vosters | Vrije Universiteit Brussel, Belgium

In this contribution, we will present a quantitative analysis of orthographic variation in the *Witnesses* corpus, a collection of Flemish witness depositions from the 18th and 19th centuries, interpreting the findings from the perspective of linguistic standardization (Rutten & Vosters, 2021).

Most studies quantifying spelling standardization as a diachronic process depart from corpus normalization and lemmatization, measuring differences between observed spelling forms and modern standard language equivalents (e.g. Tyrkkö 2020). This approach, however, presupposes a teleological model of standardization, where historical variation is reduced over time only to gradually move closer to an invariable modern standard, as part of an inevitable linear trajectory from 'dialect' to 'standard' — criticized by Watts (2011) as the 'funnel view' of standardization.

In our contribution, we will present a novel heuristic approach to quantify orthographic variation, avoiding the pitfall of teleological bias. We generate word embeddings using a pre-trained historical language model, fine-tuned on the *Witnesses* material, and combine these with Levenshtein edit distances between a particular token and its semantic nearest neighbors in embedding space, to identify orthographic variants of the same lexical item in a set of texts. We then measure the amount of orthographic variation across texts, investigating diachronic trends as well as differences between text types, scribes, and across other sociolinguistic variables.

The results of this bottom-up approach will then be interpreted in light of the standardization history of Dutch, with a particular focus on the Southern Low Countries. Although the proposed method provides us with relatively coarse-grained quantitative metrics of orthographic variation in historical texts, we will show how our findings align with earlier work on spelling change, focusing on features undergoing standardization-driven shifts in the same period.

Rutten, G., & Vosters, R. (2021). Language standardization 'from above.' In W. Ayres-Bennett & J. Bellamy (Eds.), *The Cambridge Handbook of Language Standardization* (pp. 65–92). Cambridge University Press.

Tyrkkö, J. (2020). Early modern medicine in manuscript and print: A triangulation approach to analysing spelling standardisation. *International Journal of English Studies*, 20(2), 67–93.

Watts, R. (2011). *Language myths and the history of English*. Oxford University Press.

Characteristics of early spelling errors in Arabic: The impact of visual-orthographic features

Rana Yassin & Yasmin Shalhoub-Awwad | University of Haifa, Israel

Early spelling acquisition research has traditionally focused on universal aspects such as phonology and morphology, particularly within European alphabetic writing systems. However, recently, a growing body of researchers has begun to emphasize the significance of writing system-specific dimensions and the challenges that they may pose for young spellers (e.g., Daniels & Share, 2018). The current study aimed to explore the impact of Arabic's unique visual-orthographic features, along with other universal factors, primarily phonology, on early Grade 1 spelling performance. The visual-orthographic features include (i) *the similarity of the basic letter forms*; (ii) *ligaturing*, where most of the letters within a word connect to adjacent letters; (iii) *allography*, refers to the variability of letter forms; (iv) *non-linearity*, characterized by the extensive use of diacritic-like signs (tashkil) predominantly placed above or below letters; and (v) Letter formation reflects the complexity of letter construction, where certain letters comprise several essential components necessary for accurate formation.

The study delved into spelling errors in a structured word dictation task among a representative sample of 124 Arabic-speaking first graders. Spelling errors were analyzed according to six error categories: *Visual-orthographic*, *phonological, lexical*, *distortion, orientation, and other* error categories. The results showed that while phonological errors were the most common, with around 70% of the children committing one or more phonological errors, visual-orthographic errors ranked second, with almost 50% of the sample making one or more visual-orthographic errors on the set of eight test words. These findings provide compelling evidence of the dimensions of writing system complexity in Arabic, which have been relatively overlooked in spelling research. They also highlight the significant role of visual-orthographic features in mastering basic spelling skills alongside the essential phonological knowledge required, especially in the early stages of literacy acquisition.

Daniels, P. T., & Share, D. L. (2018). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading*, *22*(1), 101–116.

How does morphological awareness before and after the onset of formal reading instruction affect later reading outcomes in a rich morphological language?

Nancy Joubran-Awadie | University of Haifa & Oranim Academic College of Education, Israel
Yasmin Shalhoub Awwad | University of Haifa, Israel

Contemporary theories and models acknowledge the significant role of morphological awareness (MA) in reading (Levesque et al., 2021; Perfetti & Stafura, 2014). However, the relationship between MA and reading may differ before and after the onset of reading instruction, owing to their reciprocal relationship. The present longitudinal study investigated the early contribution of spoken MA before and after the onset of formal reading instruction to later reading outcomes in Arabic, a rich morphological language with transparent orthography. A sample of 427 Arabic-speaking children was assessed on oral language measures (MA, phonological awareness, and vocabulary) in K and G1, reading measures (accuracy, rate, and fluency) in G1 and G3, and reading comprehension in G3. Structural Equation Modeling (SEM) was conducted after accounting for earlier word reading skills as autoregressive controls in Grade 1. Our preliminary findings have shown a high correlation between MA and vocabulary. Consequently, both constructs were integrated under the term *morpho-lexical knowledge* (MLK). Findings from SEM revealed that the nature of the spoken MLK effects differed as a function of time-point and reading outcomes. MLK in kindergarten contributed only *indirectly* to word reading accuracy and reading comprehension in Grade 3, whereas MLK in Grade 1 contributed both *directly* to word reading accuracy and reading comprehension in G3 and *indirectly* to reading comprehension. These findings suggest an early involvement of MLK that begins to emerge *before* the onset of formal reading instruction and continues to evolve, contributing more profoundly *after* the commencement of formal reading acquisition. The findings are discussed in light of the unique characteristics of the Arabic language and its orthographic system, underscoring the importance of integrating MLK in the early acquisition of reading Arabic.

Levesque, K. C., Breadmore, H. L., & Deacon, S. H. (2021). How morphology impacts reading and spelling: Advancing the role of morphology in models of literacy development. *Journal of Research in Reading, 44*(1), 10–26.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22–37.

Writing units or decades first in two digit numbers dictation tasks: The case of Arabic—an inverted writing system

Deia Ganayim | Sakhnin Academic College; Al-Qasemi Academic college; Arab Center for Mind, Brain & Behavior (ACMBB); Israel

This study investigated the effect of educational level and of the syntactic representation of numbers in Arabic writing system on the task of transcoding two-digit numbers from dictation (writing verbal number word into digits). The participants were primary, junior-high, and high school pupils and higher education students. All spoke Arabic as a mother tongue. They performed a transcoding task, namely writing two-digit numbers from dictation. Units first writing patterns-writing units first and then decades and decades first writing patterns- writing decades first and then units were collected depending on the differential syntactic structures of the two-digit number dictated (decades first: whole tens-e.g. 30, 40; units first: teen numbers-e.g. 16, 18; identical units and decades-e.g. 33, 44, remaining two-digit numbers-e.g. 23, 58). The findings reveal that in general, Arabic speakers adopt a decades-first writing pattern for two-digit numbers in the Arabic writing system, especially when it is consistent with the syntactic structure of two-digit numbers, as in whole-tens numbers. This decade first writing pattern is more evident and consistent in junior-high school, high school, and higher education than in primary school due to the improvement in mathematical skills and second and third languages. However, this pattern is modulated by the syntactic complexity of the unit–decade structure. This complexity is more pronounced in two-digit numbers whose processing is more dependent on numerical syntax. Thus, whole-tens numbers, teen numbers, and identical-decade–unit numbers are less complex than the remaining two-digit numbers. The findings of the current study stress in the Arabic writing system the contribution of text writing direction (right to left) and mathematical writing direction (left to right) to the writing of two-digit numbers to dictation in Arabic.

Ganayim, D.; Ganayim, S.; Dowker, A.; Olkun, S. Linguistic Effects on the Processing of Two-Digit Numbers. *Open J. Mod. Linguist.* **2020**, *10*, 49–69. https://doi.org/10.4236/ojml.2020.101004

Ganayim, D.; Ganayim, S.; Dowker, A.; Olkun, S. Transcoding errors of two-digit numbers from Arabic digits into verbal numbers and from verbal numbers into Arabic digits by Arab first graders. *J. Cogn. Educ. Psychol.* **2020**, *20*, 46–69

Hayek, M.; Karni, A.; Eviatar, Z. Transcoding number words by bilingual speakers of Arabic: Writing multi-digit numbers in a units-decades inverting language. *Writ. Syst. Res.* **2019**, *11*, 188–202. https://doi.org/10.1080/17586801.2020.1787298

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

Marta Guidotti | University of Hasselt, Belgium

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

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

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

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

Unger, Gerard 2006. Il Gioco della Lettura. Viterbo: Nuovi Equilibri. (Terwijl je leest, 2006)
Bessemans, Ann 2025. The automatization of rhythm deduction in type [Automatic Type Design 3]
Wilkins, Arnold J. 1995. Visual Stress. Oxford University Press: London

"The worst writing system in history": Public attitudes toward the learnability of writing systems

Dimitrios Meletis | University of Vienna, Austria

This paper investigates how 'lay' language users conceptualize the learnability – and particularly the perceived difficulty – of different writing systems. Drawing on informal online discourse from the platforms Reddit, Quora, and YouTube, it examines how linguistic knowledge, intuitive beliefs, and normative assumptions interact in shaping attitudes and ideologies toward scripts and writing systems from diverse traditions, including Hangul, Ge'ez, Thai, Mongolian, and Tangut.

The analysis is guided by four working hypotheses. First, perceptions of difficulty often reflect *cultural stereotypes* rather than graphematic structure: while Chinese is dismissed as "impossible", Greek may be jokingly reduced to "just math". Second, a *familiarity bias* suggests that systems more similar to one's own – or commonly encountered in educational contexts – are judged as 'easier'. Third, users appear to *project normative expectations* from their first acquired scripts onto unfamiliar systems, assessing them by standards such as phoneme-grapheme correspondence or (alphabetic) linearity. Finally, difficulty is frequently entangled with diverse aesthetic evaluations: Arabic may be admired as "beautiful but too hard", while Hangul is described as "clever".

The study focuses on how users art"cula'e such judgments, often invoking concepts like logic, chaos, tradition, or authenticity. Although demographic data are rarely available, self-reported details (e.g., "coming from English") allow for tentative insights into how prior literacy experience shapes perception. One interpretive framework is the *native script effect* (Gnanadesikan 2021), accounting for persistent first-script influence.

While typographic evaluations have been explored (e.g., Spitzmüller 2013), this study extends the inquiry, emphasizing how lay users assess scripts not only visually but also in terms of (linguistic) structure and logic. The paper thus contributes to our understanding of how writing systems are imagined and evaluated in the public sphere, offering a language ideological (Kroskrity 2010) perspective that complements typological and cognitive accounts with a view to an experiential dimension of graphematic and orthographic judgment.

Gnanadesikan, A. (2021). S1: The Native Script Effect. In Y. Haralambous (Ed.), *Grapholinguistics in the 21st Century 2020* (pp. 103–123). Brest: Fluxus Editions.

Kroskrity, P. V. (2010). Language ideologies – Evolving perspectives. In J. Jaspers, J.-O. Östman, & J. Verschueren (Eds.), *Society and Language Use* (pp. 192–211). Amsterdam: John Benjamins.

Spitzmüller, Jürgen. (2013). *Graphische Variation als soziale Praxis. Eine soziolinguistische Theorie skripturaler ‚Sichtbarkeit'*. Berlin/Boston: De Gruyter.

Japanese *rubi*: Layered representation of meaning

Terry Joyce | Tama University, Japan; Hisashi Masuda | Hiroshima Shudo University, Japan; Chikako Fujita | Nanzan University, Japan

In Japanese, *rubi* refers to the various graphematic conventions that underlie the materialization and deployment of marginal annotations. Undoubtedly, the most frequent form is, as in (1), where a hiragana-script rubi represents the pronunciation of a kanji-script (or mixed kanji-kana) base-word; a practice also known as furigana.

(1) 振り仮名（ふりがな） /furigana/ *kana above (or next to) kanji to indicate pronunciation*

Within the classification framework initially proposed by Joyce et al. (2023) and developed in this presentation, furigana constitutes the default type of graphematic variants—base and rubi are different graphematic representations of the same word. In contrast, the second major category is of semantic fusions—base and rubi are conjunctions of different words (or phrases). As the creative products of artfully layering different meaning representations, semantic fusions typically materialize nuanced forms of written expression, such as subtly blending stipulation and emphasis. However, given the issues of representative sampling by virtue of their ubiquity, this presentation focuses on semantic fusions within two specific contexts.

The first is the common forms of telephone-number wordplay, as mnemonics or for emphasis, that are possible because each digit in Japanese has at least one Native- (NJ), Sino- (SJ) and Foreign-Japanese (FJ) pronunciation as well as abbreviation and voicing variants. The final digits of a cram school's number in (2) are a rather contrived example, which is noteworthy for the rubi's kanji-kana-mix and its placement below.

(2) 1359 / いざ合格
Base: 1 = /ichi/ → /i/, 3 = /san/ → /za/, 5 = /go/ → /gō/, 9 = /ku/
Rubi: /iza gōkaku/ *I'll pass!* [Note: /ka/ is not represented in the base digits]

The second context is a sample spanning several months of Japan's bestselling manga anthology 週刊少年ジャンプ /shūkan shōnen jyanpu/ *Weekly Shōnen Jump*. Singling out just a couple of notable characteristics of the manga sample, (3) illustrates the frequent instances of referent-stipulation and (4) exemplifies the manga's penchant towards FJ annotations.

(3) 強盗（コイツ）
Rubi: /koitsu/ *this guy* (causal) [NJ]
Base: /gōtō/ *robber; robbery* [SJ]
→ *this robber guy*

(4) 殺し屋（プロ）
Rubi: /puro/ *professional* [FJ]
Base: /koroshiya/ *hitman* [NJ]
→ *professional hitman*

In further refining the classification framework, this presentation underscores the innovative nature of rubi as a device for the richly nuanced, layered representation of meaning.

Joyce, Terry, Masuda, Hisashi, & Fujita, Chikako. (2023). Japanese rubi: Finessing the writing/reading/meaning interface. Oral presentation given that the *writing/reading interface: 14th International Workshop on Written Language and Literacy*, 10-12 November 2023, Rome, Italy.

Plain language: A psycholinguistic approach (no forest for the trees yet)

Stefano Rastelli | LLEGS Lab: University of Pavia, Italy

Despite much discussion of plain language, only a handful of experimental studies so far (≈7–8) provide empirical evidence. I proposed *language usability* as a better notion than "easy" or "readable," defined as the extent to which sentence meaning can be not only *understood* but also *acted upon* in daily life. Using eye-tracking and self-paced reading, in my Lab we examine which linguistic features slow or facilitate the move from comprehension to action. Preliminary results show that passives are easily understood but slower to act out, gerunds always obscure argument structure, and nominalizations can help clarify who-does-what — yet such findings are still not enough to see the forest for the trees.

Profiling linguistic complexity: From human perception to language model interpretability

CNR-ILC ItaliaNLP Lab | Pisa, Italy

Linguistic complexity is a multifaceted notion, with definitions shaped by diverse frameworks such as language acquisition, typology, and computational stylometry, as well as by specific research goals. In this talk, we adopt a computational linguistics perspective and present linguistic profiling as a methodological framework for modeling complexity. We will discuss some case studies where profiling has been applied to model human perception of sentence complexity, drawing on both offline judgments and online cognitive evidence such as eye-tracking signals. Finally, we will illustrate how cognitively informed approaches grounded in reading behavior can contribute to the interpretability of large language models based on deep learning architectures, offering potential connections between human cognition and model behavior.

Reading development is about multimodal coordination and linguistic awareness: An integrative, not-so-simple view of reading

CNR-ILC Comphys Lab | Pisa, Italy

Reading aloud involves the complex interplay of visual, motor and lexical processes. However, while eye movements have been extensively investigated in the literature, much less is known about the coordination of voice, eye and finger movements in oral and finger-point reading. The talk is about how multimodal synchronized time-series of digitized reading signals can be collected and modelled computationally. Data analysis shows that coordination of eye fixations, voice articulation and finger movements is a function of the increasing sensitivity of developing readers to the linguistic structure of a written text. The way multimodal coordination changes with reading expertise is thus a fundamental marker of reading development.

Rethinking morphology-based reading in Hebrew: New findings using a finger-tracking paradigm

Naama Evanhaim, Daphna Lavi-Mudrik, & Amalia Bar-On | Tel Aviv University, Israel

Morphology is crucial for written word recognition, particularly during orthographic representation formation. This is especially evident in Semitic languages like Arabic and Hebrew, where words combine intertwined root and pattern morphemes (e.g., Hebrew תזמורת *izmóret* 'orchestra', from root *z.m.r* in pattern *tiCCóCet*). Hebrew and Arabic orthographies provide partial, opaque vowel representation, requiring readers to rely on morpho-orthographic identification. While Hebrew studies have traced identification strategies, questions remain about how different morphological forms affect reading performance. Using corpus-based pseudowords, we simulated reading of 40 pseudo-nouns across Hebrew structures varying in morphological complexity and opacity. Twenty items contrasted structures with pattern-letters at both word boundaries (e.g., **ת**דלור**ת**, like **ת**זמור**ת**) versus one boundary (e.g., דלר**ת**, like זמר**ת** *zameret* 'female singer'), paired with controls. Another twenty compared three-root-letter structures (e.g., דלר, like זמר – *zamar* 'singer'/*zemer* 'singing') with two-root-letter forms with the first or last letter replaced by a pattern letter (representing defective root allomorphy). Items were grouped by length. As vowels are partially represented, multiple vowel pattern interpretations are possible. We examined how noun-structure type influences: (1) word reading speed, (2) decoding diversity (number of vowel patterns decoded per word), and (3) morpho-orthographic identification (accordance with real morphological patterns within morphologically-based pseudowords). Pseudo-nouns embedded in short sentences were read by 40 young adults using the ReadLet platform, an innovative finger-tracking method.

We predicted a length-by-morphological-structure interaction: while longer words would be more susceptible to accuracy errors and longer reading times, this effect would be mitigated in morphologically-structured words. Conversely, shorter non-structured words would exhibit greater decoding variability. Findings will be presented and discussed from a usage-based perspective and will offer a more nuanced understanding of the contexts in which morphology facilitates word identification. While focused on Hebrew, the study's methodology and insights offer broader linguistic implications, potentially extending to other writing systems.

Unpacking the role of genre and word-level information in reading Hebrew

Galit Ben-Zvi | Ono Academic College, Israel
Amalia Bar-On | Tel Aviv University, Israel

Historically, reading research has predominantly focused on word-level processes. In recent years, however, there has been a notable shift towards examining text-level processes. Contemporary researchers recognize that while word reading skills provide a critical foundation for reading fluency, fluent text reading also requires the integration of syntactic parsing, contextual facilitation, and construction of meaning. Consequently, it is expected that texts varying in linguistic and cognitive complexity will elicit distinct reading behaviors. Empirical studies consistently demonstrate that narrative and informational texts are processed differently. Specifically, reading pace tends to be significantly slower for informational texts compared to narratives. In contrast, findings regarding word recognition accuracy are mixed; some studies suggest that when texts are appropriately leveled, students decode words accurately regardless of genre. The present study aims to investigate the interplay between word-level and contextual-level factors across different stages of reading development. Specifically, we examine whether the effect of text genre on reading performance is moderated by modulating word-level information. Hebrew provides a unique opportunity to explore this question, as children encounter two orthographic systems during reading acquisition: transparent, pointed script (full vowel representation via letters and diacritic marks), used by novice readers, and unpointed script (partial and opaque vowel representation), adopted by more advanced readers. While vowel diacritics are essential for novices, proficient readers may also benefit from diacritics for improved accuracy. However, the additional phonological information provided by diacritics may also slow reading pace. A total of 261 typically developing, native Hebrew-speaking students from 2nd, 4th, and 6th grades read four texts (two narratives and two informative texts) encompassing both pointed and unpointed versions. Text transparency and text genre influenced accuracy and reading speed, though their effects varied across grade levels, demonstrating the changing relationship between bottom-up information (diacritics) and top-down information (context) throughout reading development in Hebrew.

Kim, YG. (2015) Developmental, Component-Based Model of Reading Fluency: An Investigation of Predictors of Word-Reading Fluency, Text-Reading Fluency, and Reading Comprehension. *Reading Research Quarterly*, 50(4), pp. 459-481. doi:10.1002/rrq.107

Share, D. L., & Bar-On, A. (2018). Learning to read a Semitic abjad: The triplex model of Hebrew reading development. *Journal of Learning Disabilities*, 51(5), 444–453. https:// doi. org/ 10. 1177/ 0022219417 718198

Arabic and Hebrew: Fraternal but not identical twins

David L. Share | University of Haifa, Israel
Elinor Saiegh-Haddad | Bar-Ilan University, Israel

Arabic and Hebrew are two non-European languages written in non-alphabetic writing systems. Our presentation will survey the similarities and differences in Arabic and Hebrew in linguistic and orthographic structure and the implications for literacy learning. Both are Semitic languages, the most distinct linguistic feature of which is non-concatenated root and word-pattern morphology (McCarthy, 1981). Arabic and Hebrew are also similar in other morphological and morpho-syntactic properties, including inflectional morphology and the use of clitics, even though Arabic is much richer than Hebrew in both morphological domains, and especially so in inflectional categories. The Arabic and Hebrew writing systems feature similarities and differences too; both are abjads (i.e., consonantal writing systems) in which consonants are fully represented in writing whereas vowels are represented in a subsidiary manner partly by regular (linear) letters and partly by (optional) non-linear diacritic-like signs. Arabic is a highly cursive (ligatured) script with extensive allography; Hebrew, on the other hand, is a non-cursive script with limited allography. In both languages, children learn to read the fully vocalized (and hence phonologically transparent) version of their script but later transition to a partly vocalized version which relies heavily on the reader's lexical and morphological knowledge. The divergence between spoken and written forms ("diglossia") is prominent in in Arabic, but not Hebrew. These commonalities and differences result in patterns of literacy development and difficulties that are similar in the two languages but not identical.

The triangulation of linguistic, orthographic, and diglossic factors in reading development in Arabic

Elinor Saiegh-Haddad | Bar-Ilan University, Israel

Three conspicuous features of the Arabic language and orthography shape the development of reading in this language: (a) vowelization/vocalization, or the use of diacritical marks to represent short vowels and other features of articulation; (b) morphological structure, mainly, the predominance and transparency of derivational morphological structure in the linguistic and orthographic representation of the Arabic word; and (c) diglossia, specifically, the lexical and lexico-phonological distance between the spoken and the standard forms of the Arabic word. The talk will discuss evidence showing that the triangulation of these three features of the Arabic language, orthography and context of language acquisition and use govern the development and deployment of reading mechanisms. Moreover, developmental language and reading difficulties are better understood when evaluated within these language-specific features.

Visual languages of branding systems: The case of branding system of the *Zooba* restaurant

Aleksandra Twardokęs | University of Warsaw, Poland

In the proposed talk I will demonstrate that some approaches and methodologies used in language studies are applicable in a very strict sense in the domain of branding design.

The term "visual language" is used in the broad realm of pictorial representation. That applied among others to painting, illustration, advertisement, film. There are however examples of scientist who worked out a systematic application of language specific terms to the domains of image. Neil Cohn developed the theoretical framework to analyze comics (Cohn, 2013). He uses the term "visual language" in a very strict sense. Gunther Kress and Theo van Leeuwen propose systematic approach for image analyzes derived from linguistic. I will use a similar approach.

In a particular branding system we can find a very consequent inner structure: concrete set of graphic elements and rules, according to which these elements interact. In the design *jargon* this manual is called *key visual* and the document combing all information regarding brand's visual identity is *brand book*. By following the guideline from key visual we can design recognizable visual materials of a particular brand throughout diverse media and formats. I propose that the key visual of a particular brand and the instances of its usage (concrete graphic materials both printed and digital, for instance: business cards, envelops, outdoor advertisement, social media posts etc.) fit to a very classical distinction formulated by Ferdinand de Saussure for *langue* and *parole*. I will demonstrate this phenomena on the base of a case study: branding system of the Egyptian restaurant Zooba.

Cohn, Neil. 2013. *The visual language of comics : introduction to the structure and cognition of sequential images*. Bloomsbury Academic

Kress, Gunther, Theo van Leeuwen. 1996. *Reading Images. The Grammar of Visual Design*, Routledge

What can biscriptality offer in the realm of social media?

Rawan Abdulmonem M Almuzaini | University of Sussex, UK

The study examines the online writing practices of biscriptals, focusing on trans-scripting, a concept developed by Androutsopoulos (2015) within the framework of translanguaging. Trans-scripting is a "script-focused translanguaging" that emphasises the innovative use of biscriptal scripts and spelling; it is any "practice of representing a language in a non-canonical script in a way that is neither normatively expected nor technologically determined, but rather fluid and non-predictable" (2020, p. 304). Thus, the script serves as a meaning-making device formed by unconventional scripts and spelling rather than fixed visual systems.

This paper investigates the functions behind seven types of scripts that Saudi bilinguals can create from two primary standard scripts: Arabic and Roman, as well as from two languages: Arabic and English. These scripts are bi-scripts that involve either script-switching, code-switching, or both.

Arabic-scripted Arabic to Roman-scripted English (AA-RE)	عشاء مع view.	Dinner with view.
Roman-scripted English to Arabised English (RE-AE)	Success starts with a ستيب.	Success starts with a step.
Romanised Arabic to Arabic-scripted Arabic (RA-AA)	Yum jaded بدايات جديدة.	New day, new beginning.
Arabic-scripted Arabic to 3arabizi (AA- 3arabizi)	ويا زين ألوانه جا al5reef	Autumn is here, what a wonderful colour.
Arabised English to 3arabizi (AE-3arabizi)	كوفي تيست بيتير ويذ al97bah.	Coffee tastes better with friends.
Arabic-scripted Arabic to Arabised English (AA-AE)	يوم بيرفيكت للتمشي.	Perfect day for going out.
Roman-scripted English to Romanised Arabic (RE-RA)	Afternoon rawagan.	Afternoon relaxation.

The data were drawn from an online survey that asked 234 Saudi biscriptals about their reasons for using a mix of different non-standard scripts on social media.

The findings revealed that although most people felt that only two of the scripts were commonly used (AA-RE and AA-AE), they held opinions on all the combinations presented. The main reason for employing the AA-AE script was convenience, whereas the AA-RE script was utilised to imitate other social media users. Creativity was the predominant reasoning stated for the other scripts, and all trans-scripting practices were linked to the motivation of show-off.

Androutsopoulos, J. (2015). Networked Multilingualism: Some Language Practices on Facebook and Their Implications. *International Journal of Bilingualism*, *19*, 185–205. https://doi.org/10.1177/1367006913489198

(How) do underdeveloped writing systems distort grammar? Written Burushaski 'words' in transcription

Noboru Yoshioka | National Museum of Ethnology, Japan

Burushaski (bsk: ISO 639-3) is a language without any convention of writing, mainly spoken in northern Pakistan. Most native speakers do not write the language, and rather they employ Urdu, the national language of Pakistan, to write something in daily life (shopping memoranda, scribbling on flat walls, notifications for locals, etc.). Some individual speakers or groups have tried to transcribe the language and published a handful of books and booklets in their own transcription methods for about a half century. They are all Muslims, and so, for the most part, they want to employ Perso-Arabic letters. While they are not linguists, and so there are many problems with their writing systems. I use data mainly from the thickest Burushaski book, a translation of the Qur'an (Hunzāī (transl.) 2007), to discuss (I) repeatedly produced asymmetries in spelling; (II) bad relations between abjad systems and languages with distinctive accent; (III) problems on how to identify and treat loanwords; and (IV) grammatical errors which emerged from writing. And I will conclude my presentation as follows. All the advocates prepare special letters for Burushaski specific pronunciations which are not found in Urdu, whereas it causes the good symmetry in the Urdu spelling to become disarrayed. Burushaski has a distinctive pitch accent, but they ignore it and then their spellings are insufficient to distinguish different words. They want to write loanwords with the original spellings in the source language, provided and within that they think the words are so. The author-translator cannot write his language as he speaks, while avoiding inescapable influences from the Urdu writing system. Because he used it as the foundation of his Burushaski writing system with no notice of grammatical differences between the languages. If such underdeveloped writing systems exercised their authority, several consequent grammatical changes and reanalyses might be caused.

Keywords: Burushaski, morphological unit, transcription, Perso-Arabic abjad

Hunzāī, Ghulām ud-Dīn Ghulām [غلام الدین غلام ، ہنزائی] (transl.). 2007. *Al-Qur'ān al-Karīm: Burūshaskī tarjumā* [القرآن الکریم: بروشسکی ترجمہ ('The Noble Qur'an: Burushaski Translation')]. Gilgit: Oxford Gilgit Printers & Gilgit Kashmir Printers.

Processing letter position in transparent scripts: Insights from Korean hangul

Svetlana Alexeeva | HSE University, Russia
Alisa Lezina | Literature Translation Institute of Korea, Republic of Korea

This study investigates how the orthographic transparency and spatial configuration of Korean Hangul influence letter identification strategies during visual word recognition. Using pseudowords composed of 24 basic Korean letters arranged according to Hangul's syllabic structures (CVC-CV and CV-CVC), we created stimuli that closely resembled real words. While previous research has debated whether Hangul is processed at the syllabic or phonemic level (Kim, 2007), our focus is on letter position encoding and whether orthographic transparency alone determines processing strategies. Ktori and Pitchford (2008) thus claimed that transparent orthographies are associated with sequential grapheme processing, whereas deeper orthographies exhibit a mix of parallel and serial strategies.

Despite Hangul's high orthographic transparency—81.9% for phoneme-to-grapheme and 97.5% for grapheme-to-phoneme mappings (OTEANN model; Marjou, 2019)—a visual search task with 59 native speakers revealed signs of parallel processing. Specifically, participants identified the first letter of a second syllable as quickly as, or faster than, the final letter of the preceding syllable, regardless of structure. This absence of a slowdown contradicts a strictly sequential scanning model and supports syllable-based word recognition in Korean.

Within syllables, processing strategies varied by structure. Letters in CV syllables were identified in parallel, while CVC syllables exhibited either parallel or serial scanning depending on the presence of a following syllable. Thus, both syllabic complexity and position within a word influence letter accessibility. These findings support the view that Hangul's hybrid nature—alphabetic in components but syllabic in organization—shapes recognition strategies. Moreover, the results challenge the orthographic transparency hypothesis by demonstrating that letter position encoding interacts with script-specific spatial features.

Kim, S. Y. (2007). Visual, Lexical, and Contextual Effects on Word Identification of Korean. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 29, No. 29).

Ktori M., & Pitchford N. J. (2008). Effect of orthographic transparency on letter position encoding: A comparison of Greek and English monoscriptal and biscriptal readers. *Language and Cognitive Processes*. 23(2), 258–281.

Marjou, X. (2019). OTEANN: Estimating the transparency of orthographies with an artificial neural network. arXiv preprint arXiv:1912.13321.

Comparing the Syllabary and Logography Models of the Rongorongo Glyphic Script of Easter Island (Rapa Nui)

Gordon Berthin | University of Toronto, Canada

In the 1950s, Ventris deciphered Cretan Linear B as a syllabary encoding archaic Greek. Contemporaneously, Knorosov showed the enigmatic Mayan glyphs to be, dominantly, a syllabary (encoding the now dead Ch'olti' language). Both decipherments reinforced the conclusion that all writing systems contain significant phonetic component.

Rongorongo is comprised of approximately 300 distinct glyphs. Its 53 most frequent forms (and allographs) encompass 99.7% of the corpus[1]. Glyph frequencies approximately match the frequencies of occurrence of Rapanui language syllables[2]. Likewise, average length of linked glyphs (multi-syllable proxies) and stand-alone singles nearly equals the average syllable length of words in transcribed Rapanui[3]. Such similarities motivate the conclusion that *rongorongo* is a syllabary, perhaps including some logograms. Yet, the syllabary model has shortcomings including observed differences in distribution of single syllable glyphs vis-à-vis short words encountered in the language[4], and lack of agreement regarding which glyphs map onto which language syllables[5].

Although the glyph vocabulary of the logographic model draws significantly from Jaussen's List[6] of imperfect eyewitness recollections there's some consensus among epigraphers regarding the subjects depicted in the various glyph motifs (*viz.* fish, birds, lunar crescents). Nevertheless, logographers struggle to expand Pozdniakovs' meagre 53-glyph inventory onto core Rapanui language vocabulary. Ingenuity becomes necessary to build lexicon. Strategies include the presupposition of a telegram format to eliminate common particles[7] and application of Kizilova's[8] approximation that a given Rapanui word can represent either verb, noun, adjective, or participle. Moreover, glyphs may be multi-purposed via literary devices such as metonyms, synecdoches, rebuses, metaphors and homonyms. The resultant (logographic) model then becomes dominantly non-phonetic.

I examine the strengths and weaknesses of the syllabary and logography models for *rongorongo*. Despite its cumbrous nature, a logography seems to be most effective for emulating the *rongorongo* as it is presented upon the preserved artifacts.

Barthel, Thomas. (1958). *Grundlagen zur Entzifferung der Osterinselschrift*. Hamburg: Cram, de Gruyter.

Davis-Drake, Alan. (1989/1990). "A Layperson's Guide to Roŋoroŋo Part III." In: *Rapanui Journal* 3.4, pp. 1-2, 4-13.

De Laat, Mari (2009). *Words out of Wood*. Dulft (Netherlands): Eburon.

Horley, Paul. (2005). "Allographic Variations and Statistical Analysis of the *Rongorongo* script" In: *Rapanui Journal* 19.2, pp. 107-116.

Kiley, Erik (2025). *Deciphering Rongorongo as a Polynesian Syllabary: Linguistic and Statistical Validation*. Online: https://www.academia.edu/128174983/Deciphering_Rongorongo_as_a_Polynesian_Syllabary_Linguistic_and_Statistical_Validation?email_work_card=view-paper

Kizilova, Daria. (2016). "Translating Rapanui." In: *Multilingual* 4, pp. 51-54.

Pozdniakov, Igor and Konstantin Pozdniakov (2007). "Rapanui Writing and the Rapanui Language: Preliminary Results of a Statistical Analysis". Translated by: Sarah Turner. In: *Forum for Anthropology and Culture* (3): 3–36.

Addressing complexity in learning to read the Manchu writing system

Bai (Mark) Li| Macau University of Science and Technology, Macao

Learning to read fundamentally requires mastering orthography—acquiring written symbols and their mapping principles with sounds. This process challenges novice learners due to symbol visual complexity and the complex mapping relationship between symbols and sounds. Current reading theories largely rely on English and a few Western European languages (Share, 2025), limiting the capacity to develop a universal theory of reading. To expand the scope, we present a study on the Manchu language, a critically endangered yet under-investigated language in China with unique features of visual complexity and symbol-sound mapping complexity. Using orthographic units called *uju hergen*—symbol blocks representing phonological syllables—we investigated the role of different visual features and phonological representations in learning the Manchu orthography. Novice Manchu learners (n = 196; 89.8% female; M_{age} = 18.79) participated in two experiments designed to assess error patterns in uju hergen recognition. Experiment 1 focused on visual complexity of uju hergen and Experiment 2 on mapping complexity between Manchu phoneme marker and sound. We found a unique contribution made by connected points on uju hergen recognition. Furthermore, lower naming error rate was observed for the list of uju hergen characterised by single (one–to–one) mapping between a phoneme marker and sound compared to those by multiple (one–to–many) mapping. No switching cost was observed for the two lists presented in the blocked and mixed orders. We thus propose that, at the sub-symbol level, connected points seem to cue uju hergen symbol blocks decoding, and at the symbol level, uju hergen appear to be read by analysing their constituent phoneme markers. Findings from this understudied Manchu orthography confirm the explanatory power of the constructs of grain size, transparency, inventory size, all of them being referred to by us as orthographic scale.

Keywords: reading, Manchu, uju hergen, orthographic scale

Share, D. L. (2025). Blueprint for a universal theory of learning to read: The combinatorial model. *Reading Research Quarterly*, *60*(2), 1–51. https://doi.org/10.1002/rrq.603

Breaking written words

Jonas Romstadt | University of Bonn, Germany

Writing unfolds on a spatially bounded surface. This spatial constraint directly influences the production of written words. When writers arrive at the end of a line, they have two basic options: either the next word is shifted to the following line, or it is divided—marked by a hyphen. In both cases, spatial factors take precedence over linguistic ones initially (with list formatting as a notable exception, cf. Reißig 2015). While these processes are handled automatically in digital word processing software, handwriting remains a domain in which they must be negotiated manually. This is the starting point of the present study.

Particular attention is given to the hyphen—a punctuation mark that *divides* written words. Previous research has focused primarily on the system governing its placement (cf. Geilfuß-Wolfgang 2007; Fuhrhop & Schmidt 2014). In contrast, this study adopts a usage-based perspective: drawing on a corpus of over 1,000 handwritten German school-exams from the last 100 years, the actual hyphenation practices of writers is investigated.

The analysis reveals clear preferences in hyphen placement, which reflect the complex interplay between syllable boundaries and morpheme boundaries in written language. These patterns not only shed light on graphematic structure but also offer insights into the underlying writing processes in which they emerge. In this sense, hyphenated written words become loci of the interaction between morphology and graphematics.

Fuhrhop, Nanna & Schmidt, Karsten (2014): Die zunehmende Profilierung der Schreibsilbe in der Geschichte des Deutschen. In: *Beiträge zur Geschichte der deutschen Sprache und Literatur 136* (4), 538–568. https://doi.org/10.1515/bgsl-2014-0047

Geilfuß-Wolfgang, Jochen (2007): *Worttrennung am Zeilenende. Über die deutschen Worttrennungsregeln, ihr Erlernen in der Grundschule und das Lesen getrennter Wörter*. Tübingen: Niemeyer. https://doi.org/10.1515/9783110969351

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

Ambiguity and the creation and evolution of writing systems

Jordi Fortuny | University of Barcelona, Spain

First stages of proto-writing systems are always iconographic. For instance, the first stage of Sumerian cuneiform represents quantities and drawings of common life referents, such as animals, plants, or body parts. This system was useful for administrative purposes, but it was not complete: abstract concepts as well as grammatical categories (such as tense or case markers) could not be written. Strictly speaking, iconographic systems are not writing systems, because they do not represent speech but rather part of the physical environment.

The objective of this talk is to show that the awareness of ambiguity is a principal factor in the creation and evolution of writing systems. More precisely, it will be argued that the awareness of ambiguity was opportunistically used to increase the expressiveness of scripts, i.e., to transform proto-writing into full writing.

I will focus on rebus writing and the appearance of determinatives. On one hand, rebus writing strategically exploits homophony, and thus lexical ambiguity, in order to attain full writing. On the other hand, the use of determinatives in different scripts appears as a mechanism to reduce the amount of ambiguity created by the strategic application of rebus writing. I will provide a comprehensive comparative investigation of how rebus writing permitted, step by step, the development of full writing in the cases of Sumerian cuneiform, Egyptian hieroglyphics, and Chinese logosyllabary, as well as a systematic investigation of how determinatives are used in these scripts to reduce ambiguity.

Segmenting the relevant linguistic chunks and detecting its potential ambiguity permitted the illiterate mind to analyze speech and develop gradually more complete writing systems. Therefore, the study of writing offers a window into the awareness of linguistic knowledge. The expected results of this research line should be very relevant to the history of writing, but also to general linguistics and psycholinguistics.

Coulmas, Florian. 2003. *Writing systems. An introduction to their analysis*. Cambridge Textbooks in Linguistics. Cambridge: Cambrige University Press.

Daniels, Peter & William Bright. *The worlds writing systems*. Oxford: Oxford University Press.

Fortuny, Jordi & Lluís Payrató (eds.). 2024. *Perspectives on ambiguity*. Studia Linguistica.

Emojis and ✍️ing: What counts?

Anna Melnikova | Middlebury College, VT, USA
John David Storment | Stony Brook University, NY, USA

Emojis are ubiquitous in digital communication today. Their primary purpose is to add expressivity to written language. While emojis lack the properties of an independent writing system, they are easily incorporated into preexisting writing systems. In this presentation, we discuss semantic and morphosyntactic properties of emojis that are parallel to properties of conventional writing systems and offer an analysis of how emojis function as elements of writing cross-linguistically.

We argue that, like semantic complements in Cuneiform and Egyptian writing (Rogers 2005), emojis can add expressive information to a written utterance when they modify grammatical constituents or entire utterances (Grosz et al. 2021; Storment 2024). Moreover, just as early writing systems (e.g., Chinese, Cuneiform) utilized their symbol inventory through semantic generalizations, emojis extend their use to words with similar meanings (e.g. 🍎 'apple' can be extended to morphemes such as 'teachers' or 'New York').

Emojis may combine with orthographically represented inflectional affixes. Cases of inflectional morphology are highly restricted and vary across languages, reflecting differences in writing systems and morphological structure. For instance, in English 'she 👻ed me', the past tense suffix -ed attaches to the verb 'to ghost' (Storment 2024). In languages like Russian and Spanish, a combination of emojis with inflectional morphemes is not possible due to morphological complexity.

Emojis can also function independently in two-word utterances (e.g., 🍕⏰ 'pizza time') and modern-day rebuses (e.g., 🦁👑 'The Lion King'). While in rebuses and simple sentences emojis function as words, their classification as an independent writing system is problematic since their interpretation depends on the context and culture rather than on linguistic encoding. We ultimately show that, while emojis on their own are not enough to be considered a writing system (despite often being portrayed as such in popular culture), their status as orthographic elements is highly dependent on the amount of linguistic structure that they are embedded within.

Grosz, Patrick & Kaiser, Elsi & Pierini, Francisco. 2021. Discourse anaphoricity and first-person indexicality in emoji resolution. *Proceedings of Sinn und Bedeutung* 25, 340–357.
Rogers, Henry. 2005. *Writing Systems: A linguistic Approach.* Malden, MA: Blackwell
Storment, John David. 2024. Going ✈️ Lexicon? The linguistic status of pro-text emojis. *Glossa.*

Pictish symbols: Letting graphetics inform graphematics

David Osgarby | University of Glasgow, Scotland

'Pictish Symbols' are an early medieval symbol system attested primarily on stone monuments throughout a large area of modern-day Scotland. Despite many accounts of its claimed religious, spiritual, totemic and social significance, the system's original function remains unknown. Arguments have been advanced both that the Pictish Symbol system constitutes a narrowly defined writing system (Forsyth 1995) or else that it constitutes a non-linguistic symbol system (Sproat 2014). This paper explores whether applying a grapholinguistic framework (Meletis 2020) to the Pictish Symbol system can provide insights into its form and nature.

The majority of the Symbol corpus consists of geometric basic shapes that do not realistically depict physical entities. This paper presents evidence that geometric basic shapes are generated by transforming one of three elementary forms in one of five regular ways, and optionally combining the result with one of two modifiers: forming a 3×5×3 paradigm of basic shapes. I observe that properties of intrinsic symmetry and graphetic graphotactics provide independent support for the proposed paradigmatic model.

The geographic distribution of basic shapes and the carving methods of individual texts suggest that a systematic change occurred before c. 700 AD to the modification parameter resulting in two distinct scriptual traditions: an earlier 3×5×3 script (attested on incised monuments in the north) and a later 3×5×2 script (attested on relief-carved monuments in the south).

The structured nature of the basic shape inventory suggests that the graphetic system was designed to systematically distinguish 45 graphemes and was supplemented by 9 naturalistic animal graphemes. A cross-linguistic comparison suggests consistency with attested abugidas, such as Tagbanwa, Buhid and Haninu'o (Philippines). Furthermore, a comparison with the ogham script of the British Isles shows strong structural similarities and provides evidence of a contemporary neighbouring writing system distinguishing a 3×5 paradigm of consonant graphemes.

Forsyth, Katherine, 'Some Thoughts on Pictish Symbols as a Formal Writing System', in *The Worm, the Germ and the Thorn: Pictish and Related Studies Presented to Isabel Henderson*, ed. by Isabel Henderson, David Henry (Pinkfoot Press, 1995), pp. 85–98, http://eprints.gla.ac.uk/3447/

Meletis, Dimitrios, *The Nature of Writing. A Theory of Grapholinguistics*, Grapholinguistics and Its Applications, 3 (Fluxus Editions, 2020), doi:10.36824/2020-meletis

Sproat, Richard, 'A Statistical Comparison of Written Language and Nonlinguistic Symbol Systems', *Language*, 90.2 (2014), pp. 457–81, doi:10.1353/lan.2014.0031

Compounding patterns in Hebrew writing: A developmental study

Avital Braun, Dorit Ravid & Elitzur Dattner | Tel Aviv University, Israel

Compounds constitute a junction of lexicon, morphology and syntax, serving as testing grounds for linguistic theories of processing and development (Bauer, 2017; Libben, Gallant, & Dressler, 2021). Modern Hebrew compounds are strictly nominal, with three structural (bound, free, and double) types reflecting its historical roots, respectively associated with syntactic, semantic and discursive functions (Ravid & Zilberbuch, 2003).

The current study explored Hebrew compounding properties in developmental perspective across the school years and beyond. The study corpus consisted of 500 texts, written by native-speaking participants from middle-high SES, in five age/schooling groups – 4th, 7th, and 11th graders, young adults (aged 19-20), and older adults (aged 26-35). All 2,015 compound tokens in the corpus were classified by structure, semantic-pragmatic functions, and syntactic properties. Analyses indicated that most compounds were construct state compounds which mostly expressed complex subcategorization (hyponymic) relations as in *coffee table* or *blood test*. Most compounds were in non-subject position, which does not burden short-term memory and processing abilities. Subject compounds in post-verbal subject position were the preferred mode for introducing new topics. Texts written by younger participants had more compounds expressing genitive relations, whereas adult texts had more partitive and exocentric compounds. With age and schooling, compounds contained increasingly abstract nominals. Three major clusters were identified in the corpus: (1) the Simplex cluster, characterized by non-expanded hyponymic bound compounds, mainly prevalent in informative texts written by younger participants; (2) the Complex cluster, mostly containing syntactically expanded free and double compounds expressing genitive and partitive relations, prevalent in narratives written by older participants; and (3) the Unique cluster, containing rare compounding devices such as counting constructions, adjective-headed compounds and compounds based on proper names, associated with texts written by the older adults.

These results indicate that compounding is a central nominal device in Hebrew, relating discourse functions with morpho-syntactic and semantic compound properties, which evolves and diversifies across development al age.

Bauer, L. (2017). *Compounds and compounding* (Vol. 155). Cambridge University Press.

Libben, G., Gallant, J., & Dressler, W. U. (2021). Textual effects in compound processing: A window on words in the world. *Frontiers in Communication*, *6*, 646454.

Ravid, D. & S. Zilberbuch. (2003). The development of complex nominals in expert and non-expert writing: A comparative study. *Pragmatics and Cognition,* 11, 267-297.

How forms shape meaning: What distributional learning can and can't do in the era of LLMs

Alessandro Lenci | University of Pisa, Italy

Large Language Models (LLMs) are distributional learning systems: they acquire semantic competence from the statistical distribution of forms across massive text corpora. In this sense, LLMs represent the most advanced development of distributional semantics to date. Compared to earlier models, they appear capable of learning vastly larger amounts of knowledge, yet they still face significant limitations—many of which resemble those that characterized their predecessors. The distribution of linguistic forms is an extraordinarily rich source of information, far richer than we once imagined. At the same time, its limitations cannot be underestimated.

The Japanese language learner in the digital age: Re-evaluating the need for handwriting kanji

Niamh Kelly & Keisuke Honda | Dublin City University)

In traditional Japanese as a Foreign Language (JFL) classrooms, learners spend a considerable amount of time learning to write numerous, formally complex kanji characters (Rose, 2019). In the context of increasingly digitalised learning environments, however, the pedagogical need for handwriting kanji among undergraduate learners warrants critical re-examination (Miyahara and Fujiwara, 2020; Itō, 2021). This study explores the relevance of manual kanji writing for JFL students at Dublin City University (DCU), where learners anecdotally report devoting up to 70% of their study time to mastering kanji—despite the availability of digital tools that render the handwriting of kanji optional in many contexts.

The research investigates whether the memorisation of stroke order and composition remains necessary for learners expected to acquire in excess of 1,000 kanji characters over the course of their degree. It comprises three participant groups: 1) current first- and second-year undergraduates; 2) students spending a study-abroad year at Japanese universities; and 3) graduates working in Japanese language environments. Data collection methods include self-reported study logs, situational journals, online surveys, and focus groups.

This case study aims to identify the real-world situations in which handwriting Japanese characters is required, both during and after formal education. By incorporating perspectives from graduates actively using Japanese in professional settings, the study will assess whether the current focus on handwriting aligns with the reality of workplace demands.

Findings from this research will contribute to a broader conversation on curricular design in JFL education, particularly in light of technological advancements that have shifted the skill set necessary for language proficiency. The study seeks to inform whether handwriting kanji should remain a central component of undergraduate Japanese language instruction or whether pedagogical priorities should be rebalanced to reflect evolving communicative realities.

Itō, E. [伊藤江美]. (2021). Dankaiteki na kanji goi gakushū no tame no kyōiku mokuhyō bunruihyō no sakusei: Shakai no IT-ka o haikei to shita Nihongo gakushūsha no shoji kyōiku no ichizuke [段階的な漢字語彙学習のための教育目標分類表の作成: 社会の IT 化を背景とした日本語学習者の書字教育の位置づけ]. *JSL Kanji Gakushū Kenkyūkaishi* [JSL 漢字学習研究会誌], *13*, 20–28. https://doi.org/10.20808/jslk.13.0_20

Miyahara, A. [宮原温子], & Fujiwara, K. [藤森景子]. (2020). The needs of beginner non-native kanji learners within their kanji class [非漢字圏初級漢字学習者が漢字学習に求めるもの]. *Japanese Language Education Methods* [日本語教育方法研究会誌], *26*(2), 20–21. https://doi.org/10.19022/jlem.26.2_20

Rose, H. (2019). Unique challenges of learning to write in the Japanese writing system. In N. Yiğitoğlu & M. Reichelt (Eds.), *L2 writing beyond English* (pp. 78–94). Multilingual Matters. https://doi.org/10.21832/9781788923132-008

The internal structure of writing systems: A cognitive approach

Daniel Zagar | Université de Lorraine & CNRS (ATILF), France
Teng Guo | Royal Holloway, University of London, UK

In our view, current conceptions of reading acquisition remain incomplete and unsatisfactory. We trace the limitations of these conceptions to a prevailing assumption—that writing operates solely through a direct mapping between graphic units and linguistic units (Gelb, 1963). In response, we propose a novel account of the internal structure of writing systems that describes the cognitive mechanisms involved in their invention.

We view writing as a cognitive tool for representing speech. More specifically, it can be considered a memory aid that allows people to recall speech verbatim through material representations. Consequently, the internal structure of writing is subject to the constraints of the cognitive system, particularly to the limitations of the human memory system. Although memory readily supports paired-associate learning between print and speech, its capacity is insufficient to store every possible print-to-speech pairing individually. Faced with these limits, early writers turned to cue combination: multiple graphemes, each acting as a cue, were combined to convey a single sign— a process similar to the disambiguation strategies described by Boltz (1986) for Chinese.

In this talk, we illustrate our approach by tracing how writing developed to overcome three major obstacles:

(1) the limited usefulness of simple grapheme-to-word associations, addressed through cue-combination mechanisms;
(2) the challenge of writing polysyllabic words; and
(3) the challenge of representing complex syllables.

The systematic use of cue combination in most writing systems gradually produced distributional regularities between written and spoken forms. We conclude by discussing the implications of these regularities for theories of reading acquisition.

Boltz, W. G. (1986). Early Chinese writing. *World Archaeology*, *17*(3), 420-436.
Gelb, I. J. (1963). *A study of writing*. University of Chicago Press.

Interowriting

Alice Mazzilli | Independent artist and researcher

Interowriting is an interdisciplinary research project and artistic practice that explores writing as a process entangled with the world, not merely with the writer s cognitive state. Rooted in the theory of interoception—the perception and interpretation of signals arising from both inside and outside the body—*Interowriting* approaches writing as a dynamic extension of the body s rhythms, movements, and sensations, situating the writer in relation to the world.

Positioned at the intersection of neuroscience, philosophy, and the arts, *Interowriting* challenges the Western conception of writing as a disembodied tool of transcription of spoken language.

This approach takes form through performative and participatory methods. *Jamigraphy* is a practice of rhythmic writing guided by music and improvisation. It can take the shape of a performance or a workshop, where the artist and participants engage in large-scale gestural mark-making in response to sound stimuli. *The Walking Wall* is an artistic performance that transforms the concept of writing into a collective, embodied experience. In this piece, the artist wears a white gown and steps into the world, becoming a living, mobile wall — a moving surface that invites interaction, engagement, and spontaneous expression. Both practices highlight writing as an embodied and social process, rather than a purely linguistic or cognitive one.

Interowriting also critically addresses the symbolic erasure of handwriting in the digital era, advocating for writing as a site of resistance, presence, and human connection. It seeks to reframe handwriting not as a vanishing technical skill but as a vital and evolving mode of expression, interwoven with perception, memory, and identity.

By collapsing the boundaries between language, movement, and relation, *Interowriting* opens new ways of understanding writing as a practice that lives through the body, in time, and in space.

Visible grammatical structures in handwritten texts in German

Niklas Reinken | Leipzig University, Germany

The German writing system reflects a lot of grammatical structures, such as morphological structures or syllables. Writing shows grammar (cf. Meletis 2020: 20 f.). This is already true in printed texts, but even more so in handwritten texts, because the potential for variation is greater in those texts. Writers can more easily deviate from the usual shape of a character and use these deviations to mark grammatical patterns.

In this presentation, I will use two phenomena to show that grammatical structures in handwritten texts go far beyond those in printed texts. To this end, I use data from a handwritten corpus of 100 school-leaving exams (Reinken 2023).

I will argue, first, that gaps in the written product correlate with syllable and morpheme boundaries and, second, that inflectional reduction syllables look different from full syllables. This allows us to draw conclusions about the relationship between grammar and writing in the German writing system that cannot be drawn from a pure examination of printed scripts.

Meletis, Dimitrios (2020): The Nature of Writing. A Theory of Grapholinguistics. Brest: Fluxus.
Reinken, Niklas (2023): Die Grammatik der Handschriften. Heidelberg: Winter.

The relationship of writing beliefs to quality and understanding in STEM and non-STEM writing

Kathleen Carroll | Maynooth University, Ireland

Individual differences in writing, such as motivational factors and transactional (process focused) and transmissional (content focused) writing beliefs, have been shown to affect writing quality and discovery through writing. Research on writing quality and discovery has focused on writing beliefs, gender, writing mode and text length. However, the interaction between writing beliefs, gender and writing mode within disciplines remains under-explored in the writing quality literature. This study addresses this lacuna by examining writing beliefs as predictors of writing quality and understanding after writing in STEM and non-STEM disciplines. It also explored gender differences and mode of writing (typed versus longhand). Participants were undergraduates (N = 100; mean age = 21; female = 76) from three universities in Ireland. They completed five online questionnaires: Writing Beliefs Inventory; Writing Apprehension Scale; Short Computer Anxiety Scale; Social Media Engagement Questionnaire; General Self-Efficacy Scale. Participants wrote two short essays, typed and handwritten, in person, with order counterbalanced. Writing quality, word count and self-rated understanding after writing were measured. There were no significant differences between STEM and non-STEM across any of the measures. Transactional (but not transmissional) beliefs and word count were significant predictors of writing quality in the handwritten condition. Word count was the only predictor of quality in the typed condition. In both the handwritten and typed conditions, word count was the only predictor of understanding after writing. Implications for modality-specific approaches to writing research and instruction are discussed.

Marking the differences: What precursor skills to reading comprehension tell us

Sonali Nag | University of Oxford, UK

In this conceptual paper I reflect on two issues: a framework for assessing skills that contribute to early reading comprehension and the challenge of communicating their inherent variety. First, for an assessment framework that encompasses influential component skills to reading comprehension, attention to language-particulars must drive the effort, not what we know as the language-universals. I will demonstrate the reasons to mark the differences using the orthographic, phonological and morphological features found in the Bantu, Dravidian, Indo-Aryan and Semitic language families. The essential infrastructure for implementation of such as assessment framework is an item bank that reflects the corpus statistics of child-directed print corpora. Second, for the communication challenge, I will argue that traditional taxonomy around proto-languages hide more than they reveal about what it takes to acquire the precursor skills for reading comprehension. What is needed instead is a language grouping that captures distinctly varied learning demands. Here, the learning is specific to different tiers of the linguistic input and the demand is due to the complexity, consistency, predictability and similar other properties of the script and language. These properties shape representation and processing and are linked to the architectures of symbol sets, the sound-to-symbol and symbol-to-sound mapping principles, and multiple morphological, morphosyntactic and phonologically conditioned processes. Although much is now known about the psycholinguistics of their acquisition, an elegant grouping solution is yet to be attempted. Next steps in this endeavour are outlined.

Introduction to the Spelling Profile Assessment (SPA): A new spelling test for English primary school children

Georgia Niolaki | University of Birmingham, UK
Aris Terzopoulos | Birmingham City University

Being a competent speller and writer is key to success in school, university, and the workplace, making early intervention crucial for those with spelling difficulties. Despite the prevalence of spelling difficulties, research on spelling lags behind that on reading. Standardised tests often fail to pinpoint specific spelling challenges, and teachers/assessors acknowledge these limitations. To address this, we have developed an assessment that identifies strengths/weaknesses in phonological encoding, whole-word processing, and morphological rules in spelling.

The Spelling Profile Assessment (SPA) identifies the key processes used for spelling (phonology, orthography, and morphology) in a single battery and can be useful for teachers, SENCos, specialist assessors, and researchers alike. To the best of our knowledge, it is one of the few tests that encompasses all of these spelling components and maybe even the only one with normative comparison data for the entirety of primary. We report the psycholinguistic characteristics and standardisation of the items in the SPA, together with reliability and validity measures obtained for data from UK children in Reception year to Year 6 (*N*:1053).

Analysis indicates that the test is a valid and reliable measurement of spelling. We report single cases of individuals with phonological, lexical, and mixed deficit profiles, demonstrating the power of the SPA tool to identify particular areas of spelling difficulty. We also map the tool across the new dyslexia definition (Carroll et al., 2024).

Keywords: Spelling; Phonology, Orthography; Morphology; assessment; school-age children

Towards automated literacy diagnostics: Insights from reading and spelling error analysis

Wieke Harmsen, Roeland van Hout, Catia Cucchiarini, Helmer Strik | Radboud University Nijmegen, The Netherlands

Reading and spelling are core skills that children learn in primary school. Learning these skills is a complicated process, as it requires the contribution and interaction of many different linguistic and cognitive components, such as phonological awareness, morphological awareness, working memory, visual competence and motor skills (e.g., Caravolas et al., 2012). This results in individual differences between children in reading and spelling development. It is especially important to identify children with reading and spelling problems in due time, so that personalized support can be offered before it is too late. Therefore, standardized oral word reading (decoding) and word dictation tests are administered in schools.

These tests give standardized global reading or spelling proficiency scores that are helpful in identifying children with reading or spelling problems. However, our claim is that more and more detailed diagnostic information can be extracted from these tests that provides insight into the problems these children experience in reading and writing. The present study develops an innovative approach that produces more specific information through automatic, detailed analyses of reading and spelling errors made by Dutch third and fourth graders with reading and spelling problems. The research question we address is: To what extent do analyses of recorded and transcribed oral word reading tests and digitized handwritten word dictations provide diagnostic information on specific reading and spelling errors? These analyses are performed through an algorithm that is able to detect and classify reading and spelling errors at Phoneme-Corresponding Unit (PCU) level (Harmsen et al., 2021, 2024). A PCU is one letter or a sequence of letters that corresponds to one phoneme. We present results to answer our research question and discuss insights on automatic reading and spelling diagnostics that can be obtained through this approach. We then present our conclusions and suggest avenues for future research.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., & Hulme, C. (2012). Common Patterns of Prediction of Literacy Development in Different Alphabetic Orthographies. *Psychological Science*, *23*(6), 678-686. https://doi.org/10.1177/0956797611434536

Harmsen, W. N., Cucchiarini, C., & Strik, H. (2021). Automatic Detection and Annotation of Spelling Errors and Orthographic Properties in the Dutch BasiScript Corpus. *Computational Linguistics in the Netherlands Journal*, *11*, 281–306. https://www.clinjournal.org/clinj/article/view/140

Harmsen, W., Cucchiarini, C., van Hout, R., & Strik, H. (2024). A Joint Approach for Automatic Analysis of Reading and Writing Errors. In *Proceedings of the Second Workshop on Computation and Written Language (CAWL) at LREC-COLING 2024*, 8–17.

A dynamical framework for graphematic feature emergence: Evidence from error patterns in children's handwriting

Rosso Manuel Senesi | Independent researcher, Italy

The present work, part of a broader research program on a dynamical framework for graphematics, extends Anderson's Structural Analogy Hypothesis, focusing on distinctive graphematic features. Phonology and graphematics share the same structure-building operations within parallel production–perception loop architectures exhibiting common functional dynamics. The overall architecture draws on Kröger et al.'s (2022) speech production–perception model. We suggest that, throughout development, this network shapes a graphematic solution space (GSS) structured as a self-organizing map (SOM). Within this framework, abstract graphematic categories emerge from reciprocal interactions between graphomotor and visual state maps. Sensorimotor integration links motor plans—formalized in Task Dynamics—to their visual traces, yielding featural and "suprasegmental" graphematic structures. Given the interdependence of writing systems with other linguistic modules, the resulting GSS bidirectionally interacts with the phonological mental syllabary and mental lexicon. Through this interface, graphematic representations continually co-evolve with phonological counterparts—Kröger's mental syllabary and lexical networks—so that each system persistently recalibrates the other. We propose a dynamic extension of Dresher's (2009) Successive Division Algorithm. The GSS starts as a uniform landscape; the first visual-motor cue to become contrastive splits it into two attractor basins. Each additional cue triggers further bifurcations, successively multiplying these basins until every grapheme settles into its own stable state. The theoretical framework was employed to analyze handwriting errors produced during a copying task performed by 62 Italian children (Grades 1–7; mean age = 9.1 years) on a Wacom digitizing tablet. Error patterns were examined in relation to the developmental trajectory with respect to the progressive crystallization of the GSS. The model helps disambiguate errors arising from (i) confusion between graphemes occupying adjacent positions in the distinctive feature space, (ii) phonological interference, and (iii) gestural dynamics.

Anderson, J. (1985). Structural analogy and dependency phonology. *Acta Linguistica Hafniensia*, *19*(2), 5-44.

Dresher, B. E. (2009). *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.

Kröger, B. J., Bekolay, T., & Cao, M. (2022). On the emergence of phonological knowledge and on motor planning and motor programming in a developmental model of speech production. *Frontiers in Human Neuroscience*, *16*, 844529.

Tones as meaning making features in the orthography of Grassfields Bantu language spoken in Cameroon: The Case of Fe'efe'e

Djomeni Gabriel Delmon | University of Dschang, Cameroon

One of the key challenges in analysing Grassfields languages is the exploration of their tonal patterns and their multiple functions in those languages. These challenges met in language analysis also influence the writing systems of such languages. This study particularly focuses on Fe'efe'e, a Bamileke Grassfields-East language spoken in the Southernmost part of the Bamileke area of Cameroon. Out of the controversy surrounding tone marking in the orthography of Cameroon tone languages, it builds on natural data collected from gradual fieldwork experiences coupled with classroom practices and group discussion to argue that tone marking in the orthography of Fe'efe'e has become a daily practice in the written form of the language. The practice has been strengthened and generalised with the introduction of Cameroon languages in formal education, and based on the General Alphabet of Cameroon Languages (GACL). Although the importance of tone marking in orthography has already been stressed by some scholars and has been combatted by others, notably those working on tone languages across Bantoid and Bantu languages and even non-African languages, we pay special attention on Fe'efe'e in this study to unveil what peculiarities the language exhibits with regard to tone behaviour in orthography. The proposal therefore questions the relevance of tone marking in the orthography of tone languages such as Fe'efe'e and the validity of the controversy around the issue. Building on traditional linguistic theoretical basis and autoethnography, the analysis exposes that at the level of the lexicon and grammar of the language, tones play both a distinctive and a grammatical function as already well-known in the literature. Hence, they can be associative markers coupled with segmental morphemes, part of tense, aspect or mood markers for some tenses. The distinctive and grammatical functions play a semantic role very relevant and useful in the orthography of the language. Though this is not new, illustrations from the Fe'efe'e language help expand knowledge on the relevance of tone marking in the orthography of Grassfields languages and in formal classroom and literacy teachings to ease the pace of acquisition of orthography principles.

Keywords: tones, Fe'efe'e, grammar, orthography, function