The internal structure of writing systems: A cognitive approach

Daniel Zagar
Université de Lorraine
& CNRS (ATILF), France
daniel.zagar@univ-lorraine.fr

Teng Guo
Royal Holloway,
University of London, UK
Teng.Guo@rhul.ac.uk

Talk presented at the 15th AWLL international workshop on writing systems and literacy @Pisa, Italy
31 October 2025

Classical conception of writing: Mapping assumption

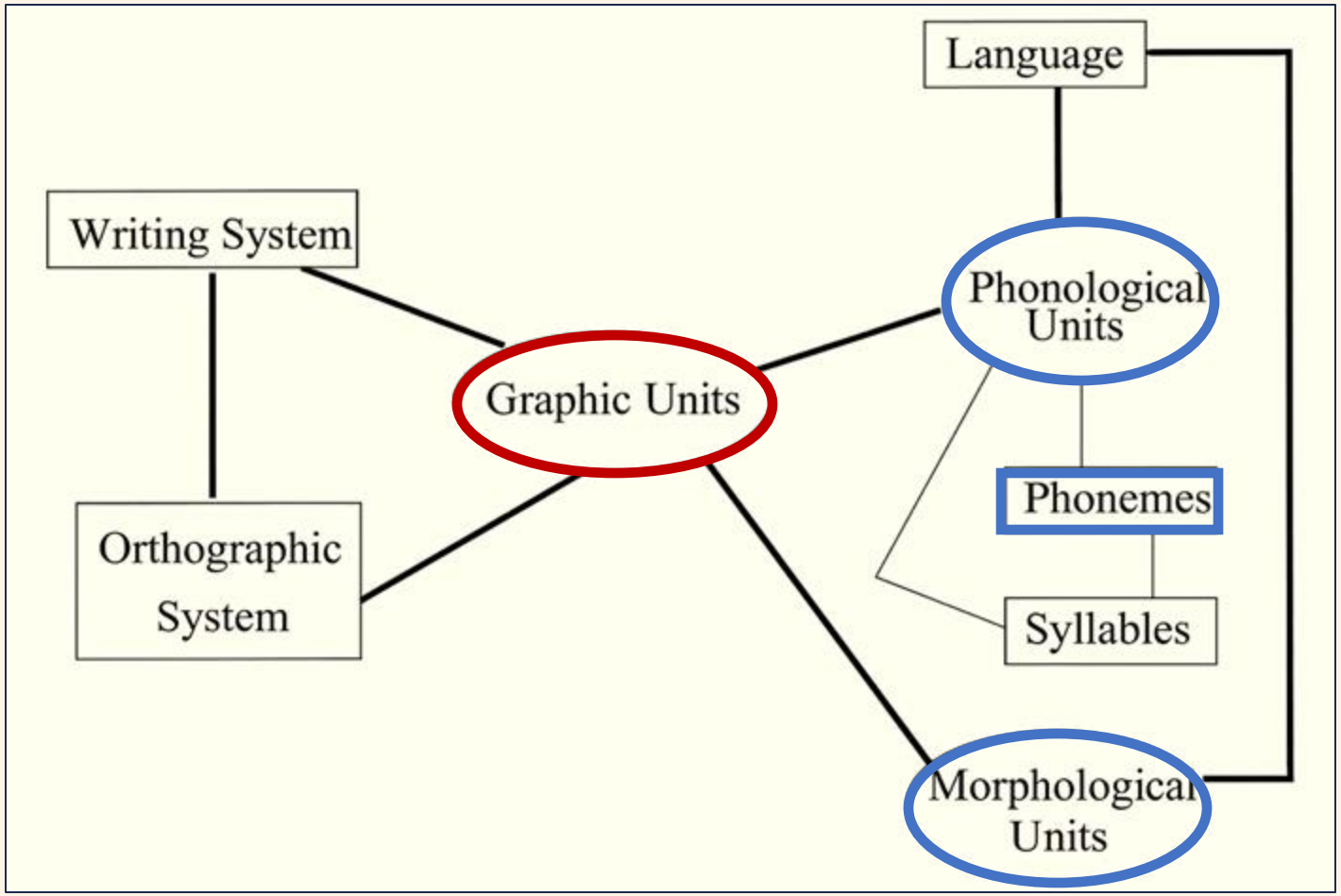
Linguistic Element ⟷	*Written Sign*	*System of Signs*
Single Sound (phoneme)	Letter or Alphabetic Sign	Alphabet or Alphabetic Writing.
Syllable	Syllabogram or Syllabic Sign	Syllabary or Syllabic Writing.
Word	Logogram or Word Sign	Logography or Word Writing.
[Phrase	Phraseogram or Phrase Sign	Phraseography or Phrase Writing]
[Prosodic Feature	Prosodic Sign or Mark	Prosodic Writing]

p.14, Gelb, 1963

- No two scholars agree completely on how writing systems should be classified (e.g., Gelb, 1963; Sampson, 1985; De Francis, 1989; Daniels, 1996; Rogers, 2005; Gnanadesikan, 2009)

- For a discussion of this phenomenon, see Joyce & Borgwaldt. 2011; Joyce 2016; Joyce & Meletis, 2021.

This conception of writing strongly influences theories of reading acquisition.



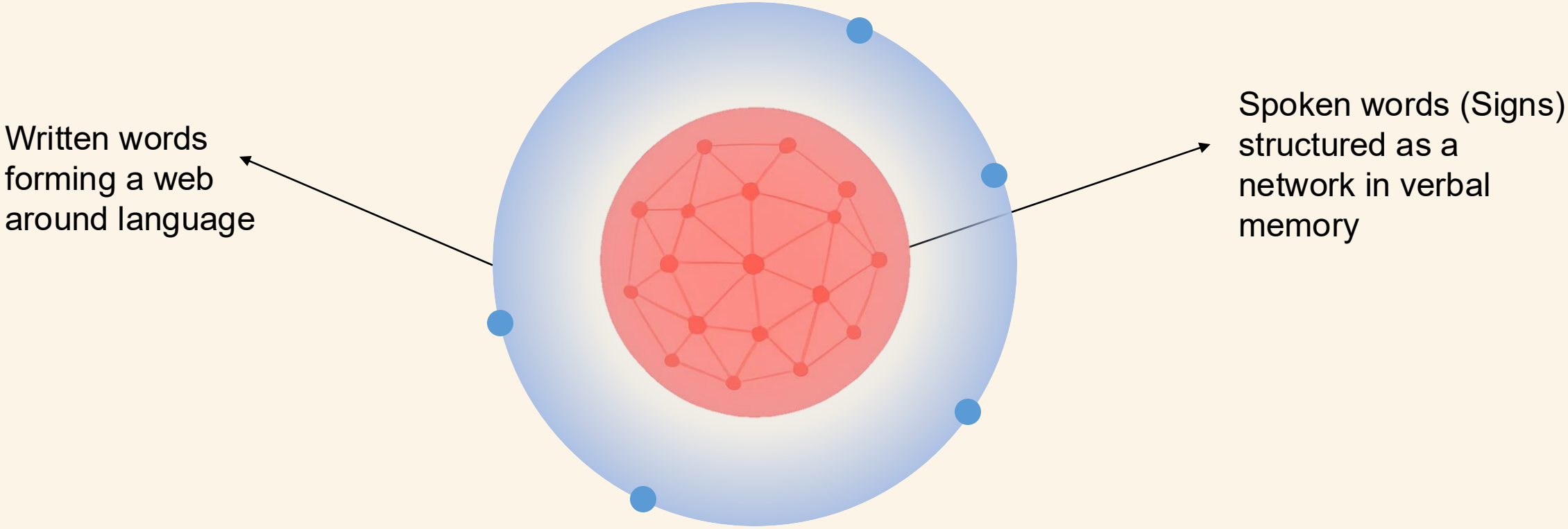
p.32, Rayner et al., 2001
(1628 citations up to 26 October 2025, from *Google Scholar*)

- Writing systems based on different mappings principles
- Learning to read = understanding how these mappings work
- In alphabetic systems: the alphabetic principle
  → letters ↔ phonemes

Conception of writing: A cognitive perspective

A set of writing units (graphs) serves as cues that enable access to verbal memory, leading to the precise and generally unambiguous activation of a spoken word (or a Sign) in the language.

Writing is viewed as a sophisticated memory aid, allowing the retrieval of each word in the language precisely from a limited set of graphic cues.



Plan

I. The invention of writing

II. The spread of writing

III. The French orthography

Discussion: How our cognitive perspective influences theories of reading acquisition

I. Invention of writing: Logography

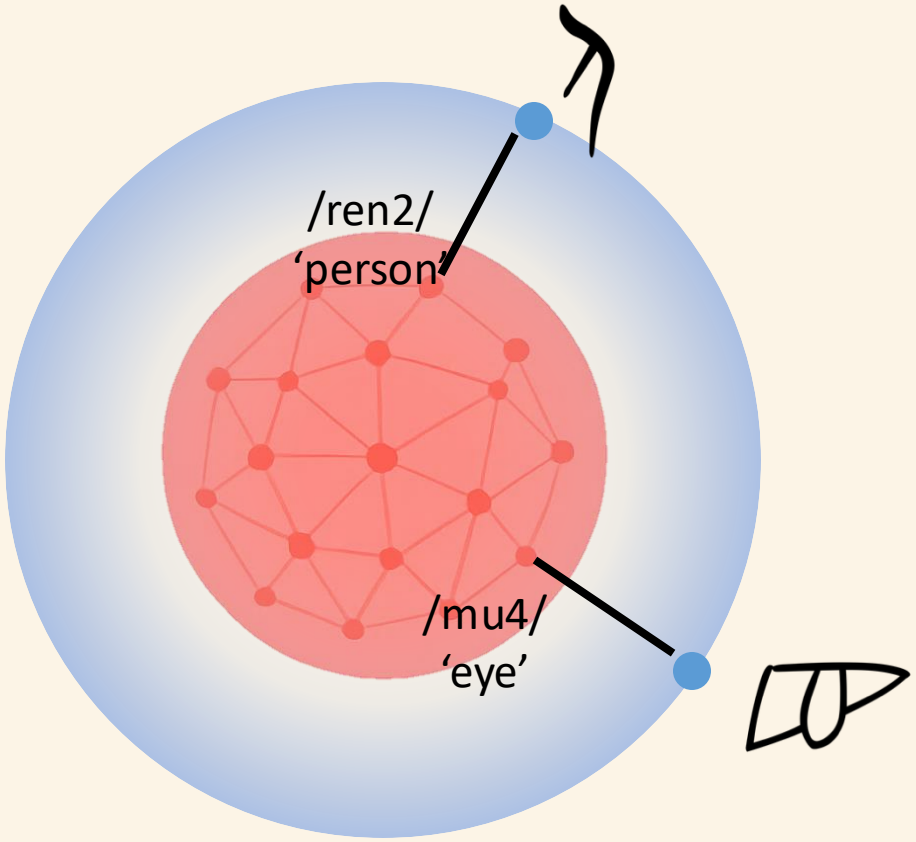
Each graph corresponds to a spoken word in the language.

Cognitive functions:

- Provides a first access to verbal memory.
- Written word = tangible, stable object
- Represents both a meaning and a sound → *Saussurean Sign*
- Reveals language productivity → combining units allows full expression (*double articulation*, Martinet 1970)

However,

- Expansion = more direct graph—word associations
- Cognitive limit ≈ a few thousand words



How can new words be written in a way that keeps their association with spoken words manageable for human memory?

I. Invention of writing: Two techniques

Creating something new from what already exists.

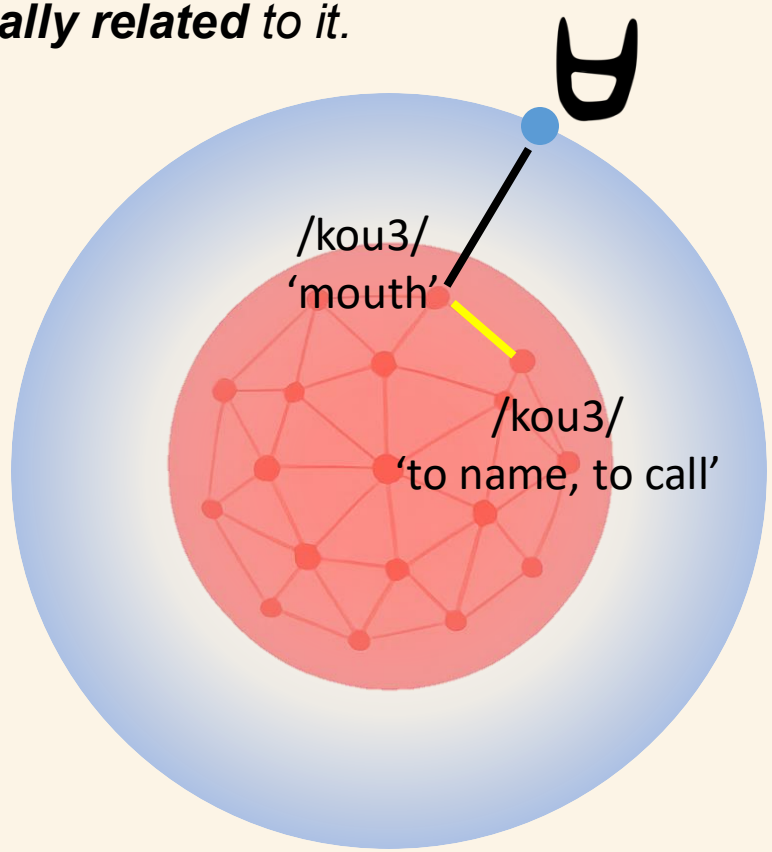
Evidence from the earliest writing systems shows the use of **two techniques** to expand the written vocabulary. A new word could be written by:

(i) **extending** the use of an already existing graph, and

(ii) **combining** two existing graphs.

I. Invention of writing: Semantic extension

*Using an existing graph to represent a word that is **semantically related** to it.*



Cognitive logics:

- Relies semantic associations between words within verbal memory
- Existing written words act as cues to retrieve new words from verbal memory

However,

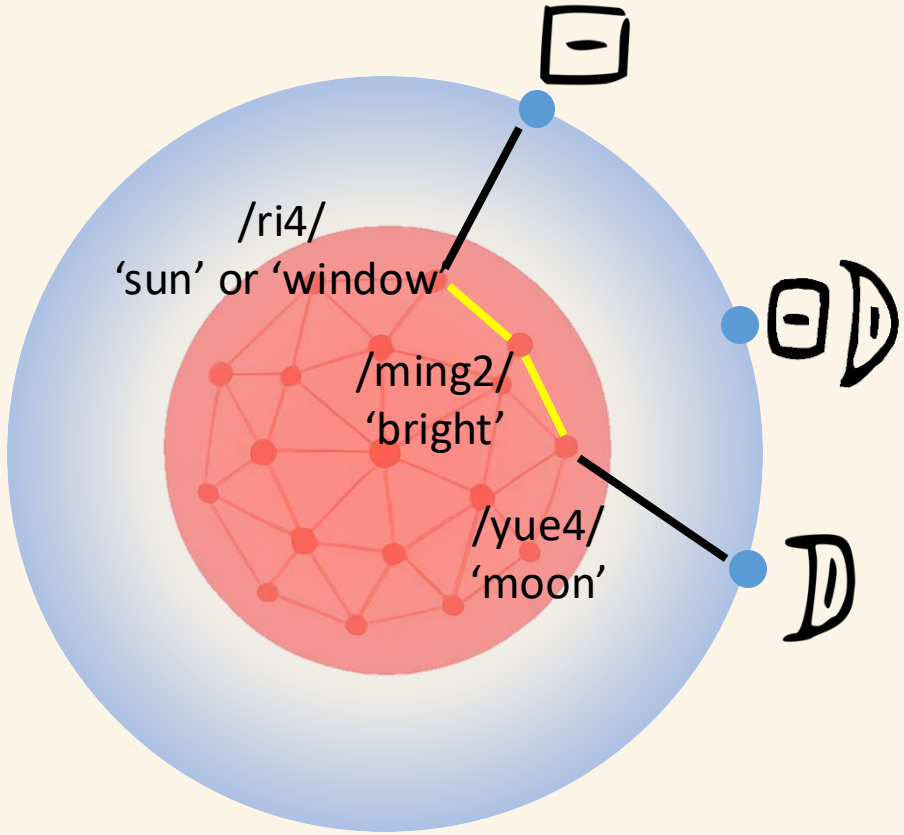
extensive use of this technique could lead to ambiguity in writing, due to homography — one written word being associated with two or more different spoken words.

I. Invention of writing: Semantic combination

Combining two or more graphs for their semantic value to form a new word related to the meanings of the original graphs.

Cognitive logics:

- Relies on semantic associations between words within verbal memory
- Each graph activating a distinct sign, combining their meanings (signified) to denote a new word
- Reduces ambiguity and resolves the problem of homography



However,

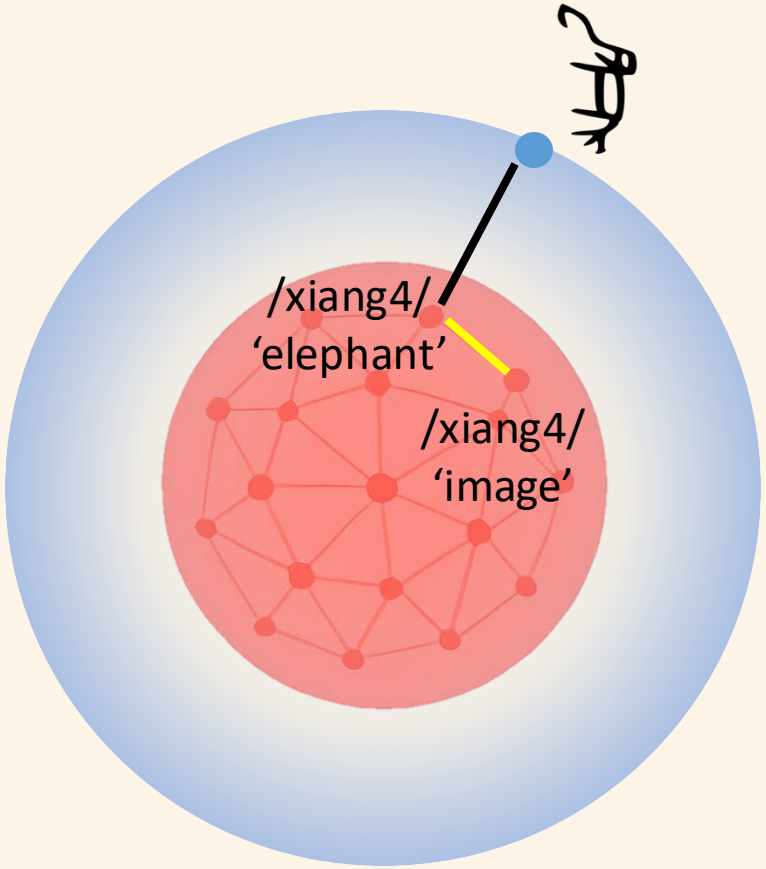
semantic combinations are highly context-dependent (e.g., cultural, geographical, or even individual), making new words difficult for others to interpret.

I. Invention of writing: Phonological extension

Using an existing graph to represent another word whose pronunciation is identical or similar to the original.

Cognitive logics:

- Relies on phonological associations between words within verbal memory
- Existing written words act as cues to retrieve new words from verbal memory



Note. This process is described as the principle of phoneticization (Gelb, 1963; De Francis, 1989) and as the paronomastic use (Boltz, 1985).

I. Invention of writing: Interim summary

- Logography enabled concrete manipulation of written words.
- Logography led to the discovery of key properties of the human language system:
  - productivity of the first articulation
  - duality of Saussurean signs → meaning + sound
- Newly created words from the logographic vocabulary through extension and combination reflect attempts to use verbal memory as a network onto which written words could be anchored.

However, none of the processes we have just described made it possible to create the structure of a full writing system.

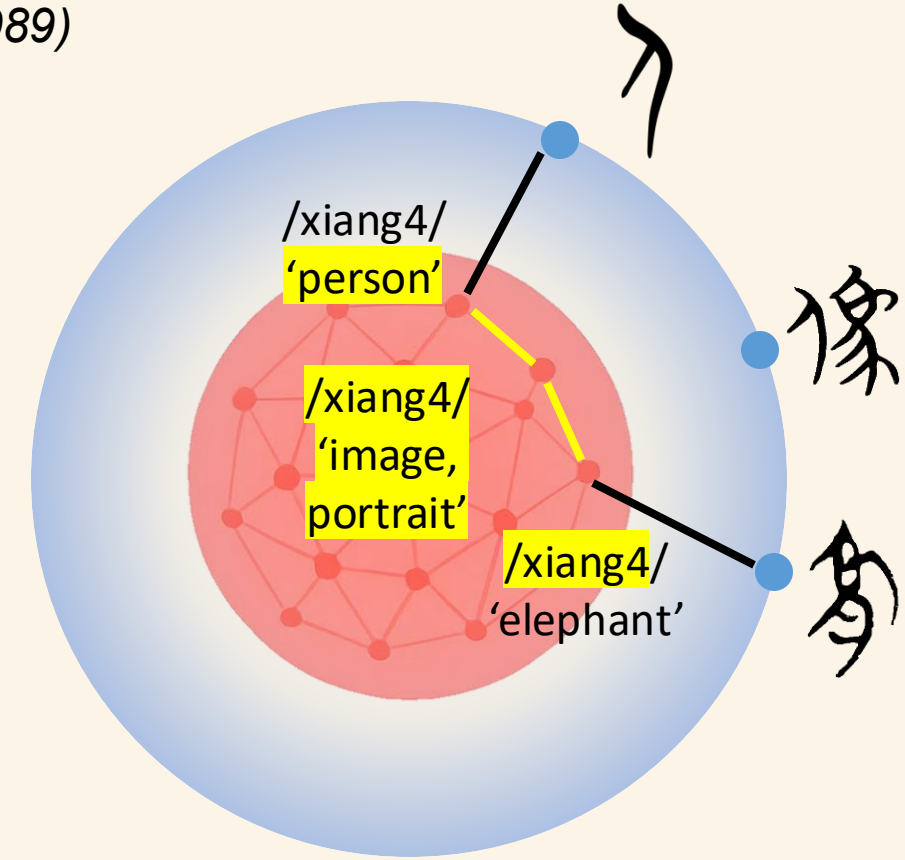
Although none of these processes alone allowed for the writing of all the words in a language, all the necessary tools were already available to construct a full writing system.

I. Invention of full writing: Combination of phonological and semantic extensions

Combining two types of graphs: one providing semantic information and the other phonological information (Boltz, 1986; the SP combination, De Francis, 1989)

Cognitive logics:

- Relies on phonological and semantic extensions between words within verbal memory
- Existing graphs act as cues to retrieve new words from verbal memory
- Phonological cues reduce ambiguity from homography through precise phonological information
- Semantic cues maintain the creative and open-ended nature of the semantic cue



This process makes it possible to write all words in the language with a limited number of graphs, staying within human memory constraints.

I. Invention of writing: Summary

- Full writing was invented through the combination of graphs that served as anchors to verbal memory.
- It emerged from the combination of two cognitive processes:
  - Semantic extension
  - Phonological extension

- As it developed, Chinese writing made use of two cognitive properties:
  - Perceptual (or reading) units
  - Distributional (quasi) regularities

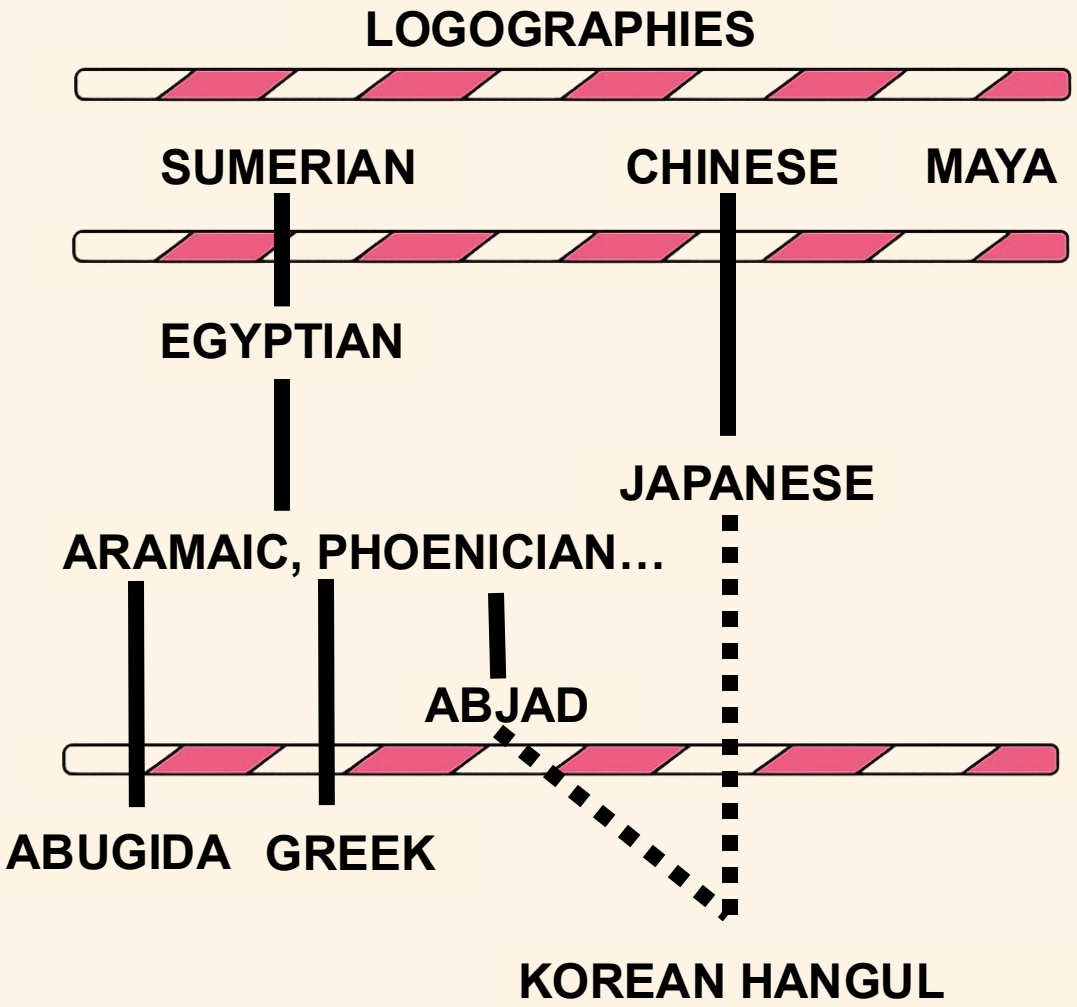
semantic	Phonetic 264 敖 (*áo*)	Phonetic 282 參 (cān)	Phonetic 391 堯 (*yāo*)	Phonetic 597 甫 (*fǔ*)
9 亻 'person'	傲 (*ào*: 'proud')	傪 (*cān*: 'good')	僥 (*jiǎo*: 'lucky')	俌 (*fù*: 'help')
64 扌 'hand'	摋 (*ào*: 'shake')	摻 (*shán*: 'seize')	撓 (*nǎo*: 'scratch')	捕 (*bǔ*: 'catch')
75 木 'wood'	槸 (*āo*: 'barge')	槮 (*shēn*: 'beam')	橈 (*náo*: 'oar')	桶 (*fú*: 'trellis')
85 氵 'water'	滶 (*ào*: 'stream')	滲 (*shèn*: 'leak')	澆 (*jiāo*: 'sprinkle')	浦 (*pǔ*: 'creek')

p.107, De Francis 1989

Why didn't other peoples simply adopt the structure of Sumerian or Chinese writings?

II. The spread of writing: Three barriers

Grammatogenies



The barrier of memory capacity

The barrier posed by polysyllabic words

The barrier of syllabic complexity

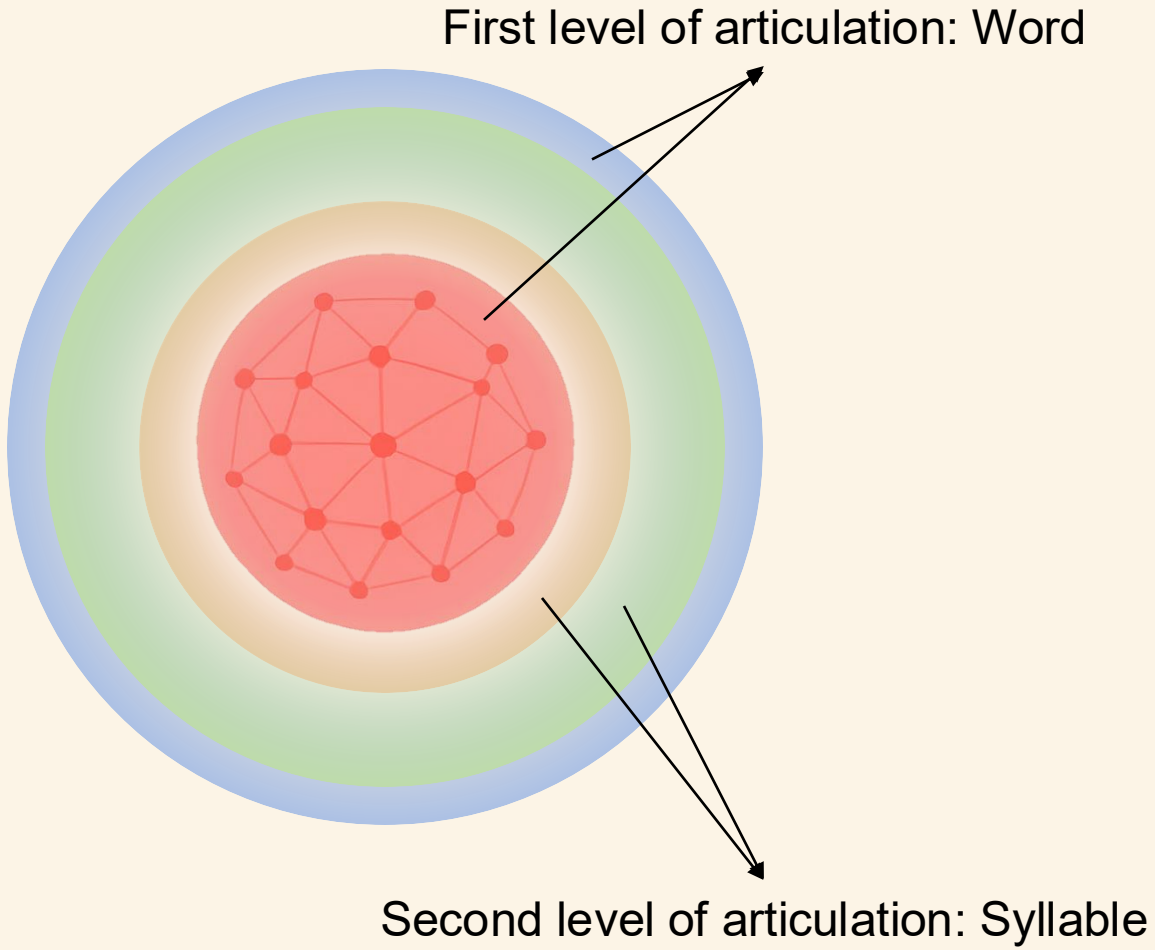
II. The spread of writing: Solution for polysyllabic words

Barrier posed by polysyllabic words:

- The cognitive mechanisms of extension and combination relied on homophony, which only highly monosyllabic languages (like Sumerian or Chinese) possess in sufficient quantity.
- Polysyllabic languages lack sufficient homophones: their words do not cover all syllables in the phonological space.

Solution:

- Speakers decomposed words into syllabic units, each of which could be represented by a graph.
- This introduced a second level of articulation — the syllable — as an intermediate representational layer between speech and writing.



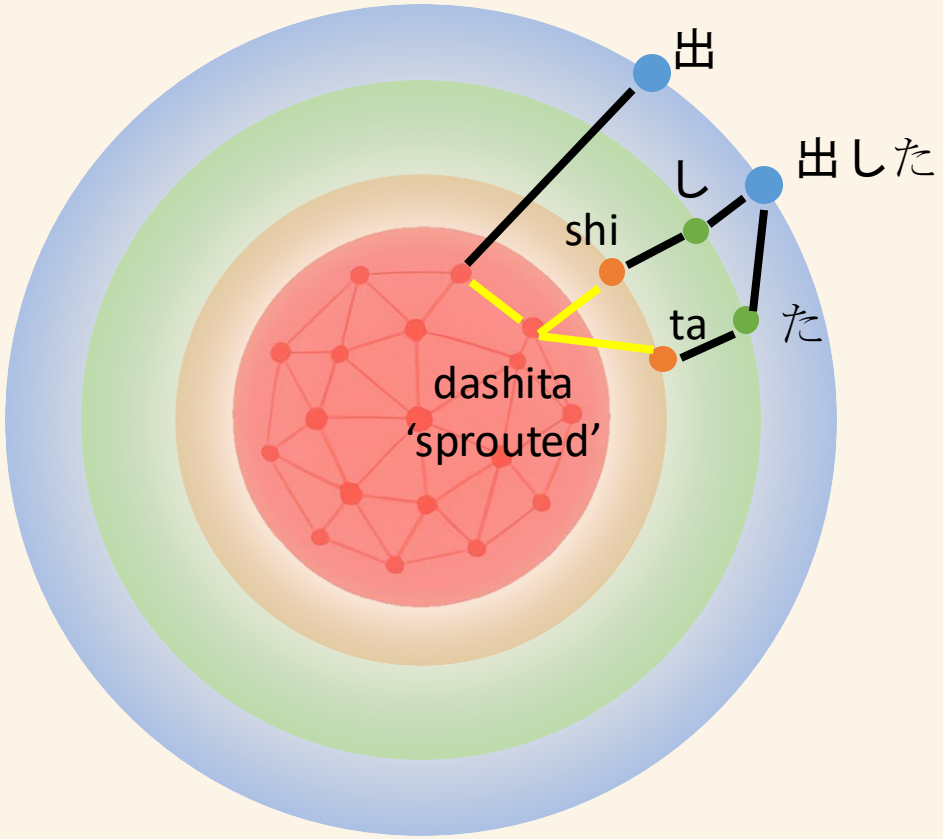
II. The spread of writing: The path of Japanese

コスモスの種が芽を出した

Cosmos's seed sprouted a bud.

Using the second level of syllable (orange and green layers):

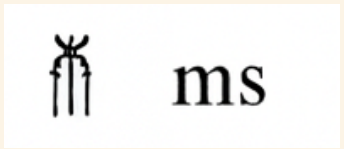
- Restored productivity (ability to write all words with limited graphs).
- But made word access slower and less direct, since reading required recombining syllables to reconstruct words.



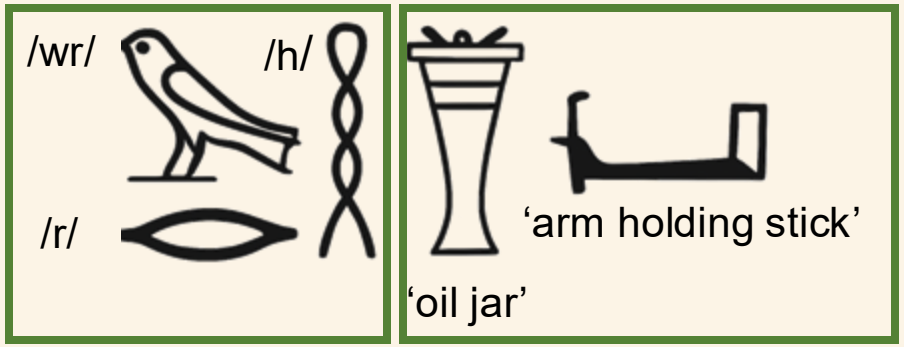
II. The spread of writing: The path of Egyptian

- Adopted combination of semantic + phonological extensions

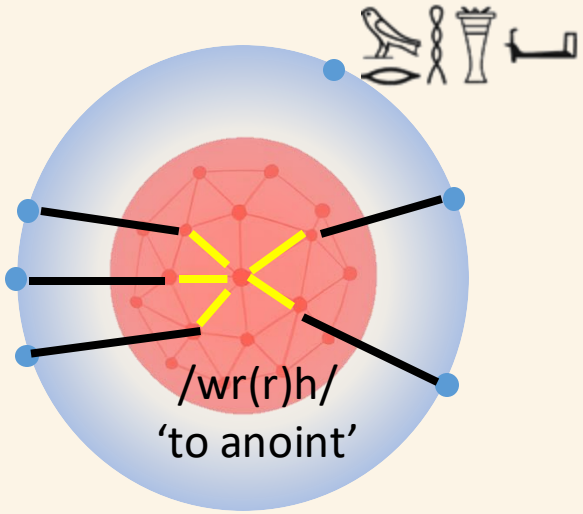
- How was phonological extension represented in this polysyllabic language?
  - Homophony needed
  - Quasi-homophony based on sequences of syllable onsets (three, two, or one consonant), differing only by vowel sounds



Ritner, 1996

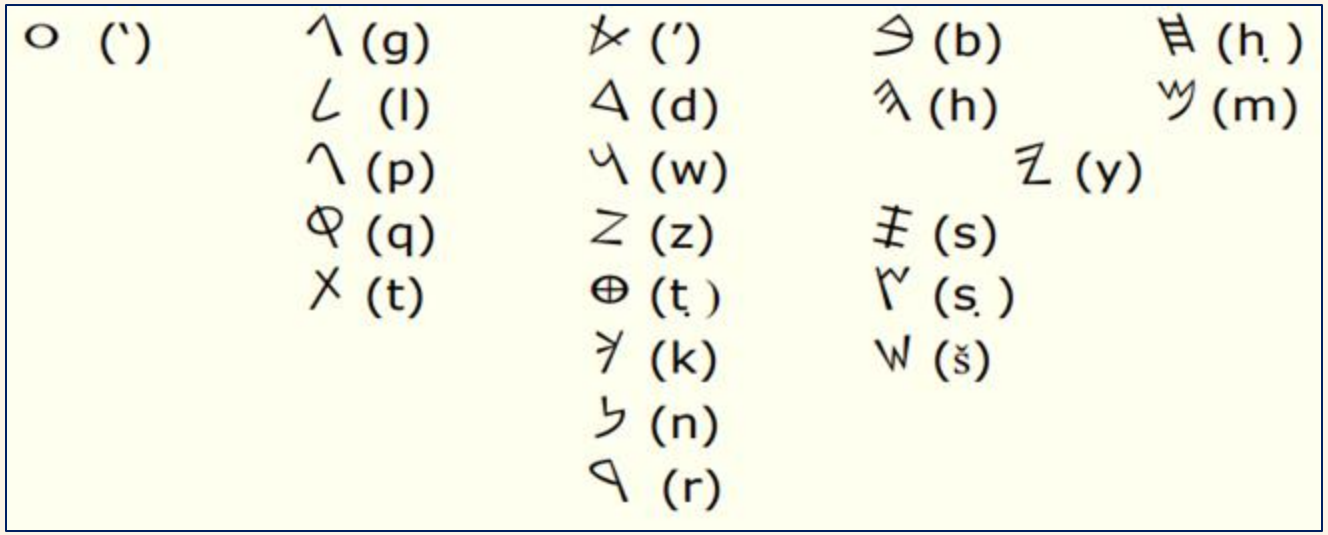


/wr(r)h/
'to anoint'



II. The spread of writing: Western Semitic simplification

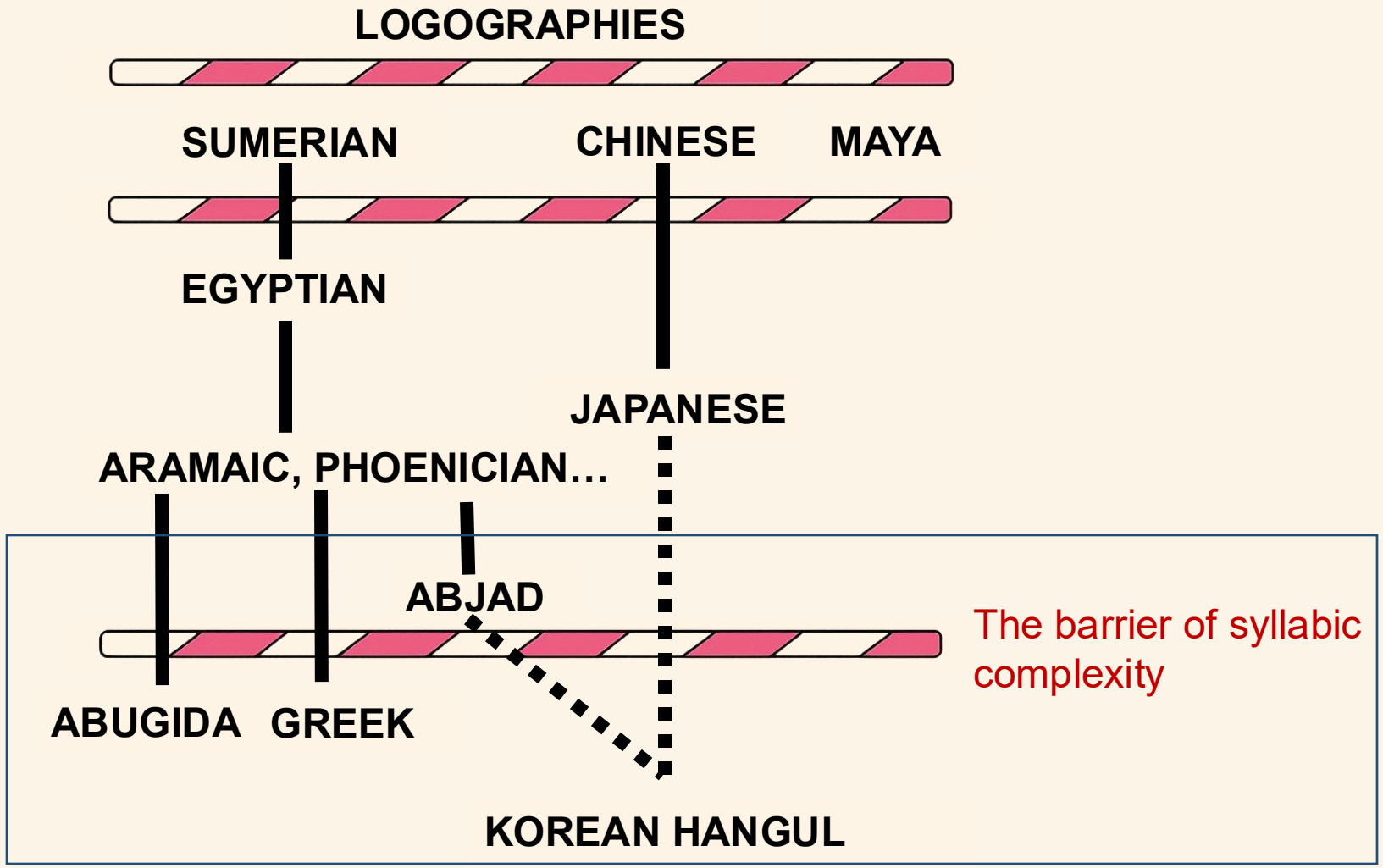
- Simplified the Egyptian system → used graphs for a single syllable onset
- Graphs represented consonants or syllables with unspecified vowels (Gelb, 1974)
- Functionally:
  - conveyed the syllable onset (isolated)
  - represented the word root (through three consonantal graphs)



Phoenician

This simplification ultimately led to the development of Arabic and Hebrew.

II. The spread of writing: Three barriers



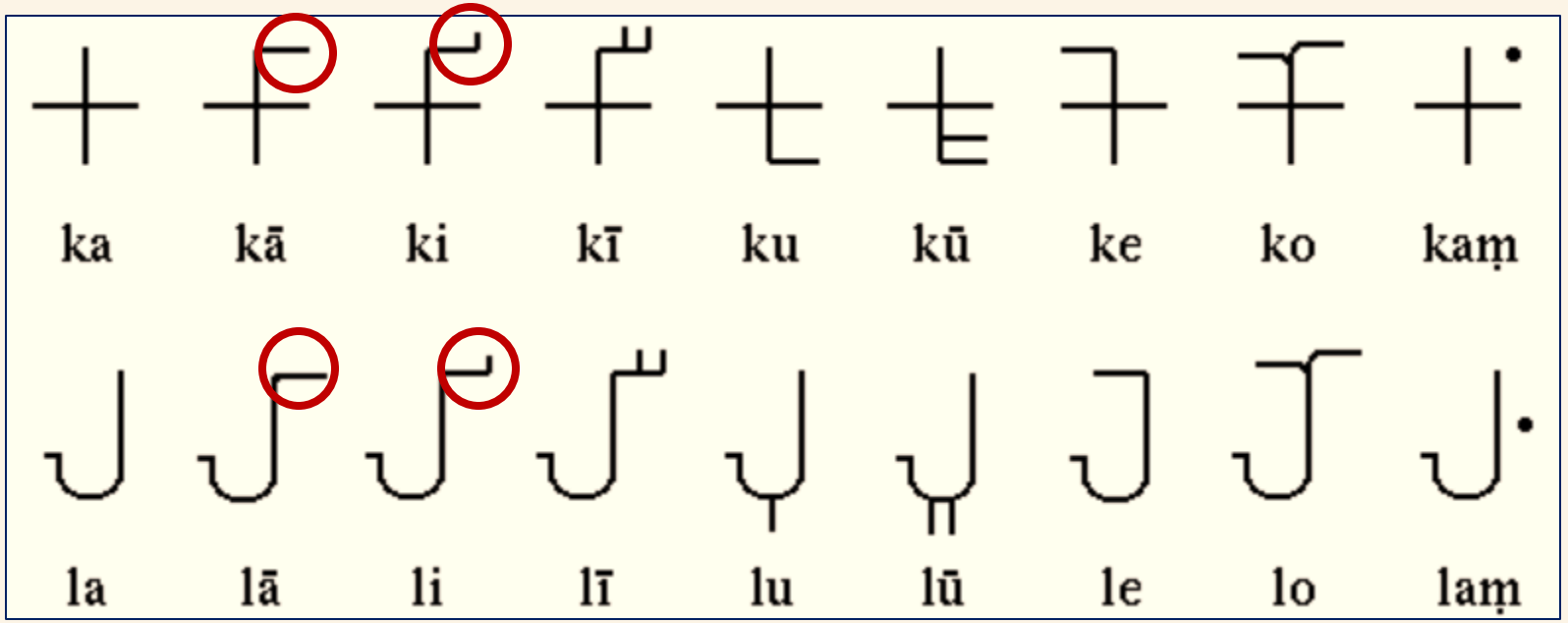
The barrier of syllabic complexity

Languages with ‘complex’ syllables
- More than 3 vowels
- Complex onsets (CC-, CCC-…)
- Complex coda (-VC, -VCC…)

Barrier: A single graph is not sufficient to denote the syllable, since homography would be too high.

II. The spread of writing: Solution for syllabic complexity

Writing complex syllables requires additional cues for accurate representation; each syllable is represented by a combination of graphs in addition to the syllable onset.



Brahmi writing as example

II. The spread of writing: Solution for syllabic complexity

Writing complex syllables requires additional information; each syllable is represented by a combination of graphs in addition to the syllable onset.

Greek takes a different graphic engineering approach:

θν

ββρ

Unclear syllables

αθήνα

βαρβαρός

Specify the identity of the syllable by combining the consonantal onset marker with new or existing graph, like “diacritics written in a row.”

Αθήνα

βαρβαρός

Additional graphs were allocated the same visual space as consonantal onsets → linear alphabetic writing

II. A modern invention —Korean Hangul

- Loss of syllable boundaries in alphabetic writing
- Hangul is not written linearly but arranges its letters (jamos) into syllable blocks.
- Preserves both syllable perception and alphabetic principles

읽기 학습은근본적인 학습입니다.

(ilg-gi hak-seu-beun geun-bon-jeo-gin hak-seum-im-ni-da)

‘Learning to read is a fundamental form of learning.’

III. The French orthography

Since Greek, alphabets are writing systems in which the second level of articulation — syllable is represented through a combination of graphs (originally syllable onsets).

The way graphs are combined to represent syllables is called *orthography*. There are three types of orthography (combinations of graphs):

- Vocalic syllables
- Syllables in monosyllabic words
- Syllables in polysyllabic words

Vocalic syllables

- Problem: only 5 vocalic letters and more than 10 phonological vowels in French
- Solution in French: combining existing graphs, e.g., /ã/ → AN, AM, EN, EM; /u/ → OU

Syllables in monosyllabic words

- Problem: to avoid homography that may be caused by homophony
- Solution: to add graphs (as in Chinese or Sumerian)
  - *de, deux*
  - *cent, sans, sang, sent*

Syllables in polysyllabic words: quasi-regular use of graphs → distributional regularity

	B	C	D	F
A	/ba/	/ka/	/da/	/fa/
E	/bə/	/sə/	/də/	/fə/
I	/bi/	/si/	/di/	/fi/
O	/bo/	/ko/	/do/	/fo/
U	/by/	/ky/	/dy/	/fy/

Discussion: How our cognitive perspective influences theories of reading acquisition

- The construction of French orthography relies on (at least) two mechanisms: transitional regularity and distributional regularity.
- These mechanisms draw on the brain's statistical learning abilities.
- The existence of these two types of statistical learning has been demonstrated in prereaders

Theoretical framework
Guo et al., 2023

Transitional regularities:
Doignon-Camus & Zagar, 2014

Distributional regularities:
Vazeux et al., 2020

Thank you!