Visual Complexity of *Akṣaras* and Orthographic Neighbors

Observations Based on the Telugu Writing System

Vasanta Duggirala, Surampudi Bapi Raju & Soham Korade

Abstract. The present study was designed (1) to examine aspects of graph and grapheme complexity associated with Telugu *akṣaras* in the script that is currently in use, and (2) to estimate quantitatively, the impact of grapheme complexity on the Orthographic Neighborhood (ON) size in 200 two-syllabic words that were part of the 1000 most frequently occurring printed words listed in a Telugu-text corpus. For all the graphs and graphemes in the Telugu script, four aspects of their visual complexity, viz., Perimetric Complexity (PC), Disconnected Components (DC), Connected Points (CP), and Simple Features (SF) were measured. With the help of an algorithm, ONs for the 200 bi-syllabic words were generated by substituting entire *akṣara* or by altering its sub-components without changing the position of the *akṣara* in the target word. The lexical status of each ON was verified using an online dictionary. Statistical analysis of the results revealed that the complexity values for the PC and DC components were strongly negatively correlated with ON, suggesting that greater visual complexity of graphemes is associated with fewer ONs in Telugu. The implications of these results are discussed briefly, and some topics for future research are mentioned in this paper.

1. Introduction

Considerable research on writing systems reported to date has been driven by the idea that learning to read in any script depends primarily on how the minimal units of writing (graphemes) map onto linguistic units such as phonemes, syllables, and morphemes. This idea about

Vasanta Duggirala 0009-0005-2139-3581
Retired Professor of Linguistics, Osmania University, Hyderabad, India
E-mail: vasantad@gmail.com

Surampudi Bapi Raju 0000-0003-2204-0890
Cognitive Science Lab, International Institute of Information Technology (IIIT), Hyderabad, India
E-mail: raju.bapi@iiit.ac.in

Soham Korade 0009-0005-2518-0313
Student of Engineering, IIIT, Hyderabad, India
E-mail: soham.korade@students.iiit.ac.in

Y. Haralambous (Ed.), *Grapholinguistics in the 21st Century 2024. Proceedings*
Grapholinguistics and Its Applications (ISSN: 2681-8566, e-ISSN: 2534-5192), Vol. 12.
Fluxus Editions, Brest, 2026, pp. 443–468. https://doi.org/10.36824/2024-graf-dugg
ISBN: 978-2-487055-12-4, e-ISBN: 978-2-487055-13-1

representational mapping informed existing major typologies of writing systems, viz., alphabets (representing both vowel and consonant phonemes as in English), abjads (consonants only as in Hebrew), alphasyllabaries/abugidas (*akṣaras* as in Hindi), syllabaries (syllables /*morās* as in Japanese), and morphosyllabaries (morphemes as in Chinese). The terms for the typologies of writing systems were directly applied to all scripts in each type. Thus, scripts associated with all Brahmi-derived Indic languages, including Telugu, were referred to as alphasyllabaries or abugidas since the minimal orthographic unit across all these scripts is *akṣara*, an orthographic CV syllable with an inherent vowel that is assumed to have similar functional value. Details about the origin of Brahmi-derived writing systems and the cultural contexts in which scripts associated with an abugida and an abjad can come into contact in India are provided by Ramanujan and Weeks (2019).

The dominant discourse around the principles of representational mapping also animated much published psycholinguistic research on how reading takes place in *akṣara*-based writing systems (see, e.g., Nag (2014; 2017); Padakannaya, Pande, Saligram, and Shruthi (2016) for Kannada; Bhuvaneshwari and Padakannaya (2014) for Tamil; Duggirala (2019) for Telugu; Chatterjee-Singh and Sumathi (2019) for Hindi; Nambiar, Kishore, and Bhargava (2023) for Malayalam).

In recent years, there has been a realization that the multilingual, multiscriptal ethos of India poses many challenges to existing discussions of representational mapping. Some of these challenges include: (1) the arrangement of *akṣara* and its components in writing (most Indic scripts have a non-linear, hierarchical arrangement of akṣara's components around its base unlike letters in an alphabet) with the degree of non-linearity varying across different Indic scripts (2) differences in the extent of mapping between phonological features and orthographic units, and (3) level of graphic complexity of *akṣaras* and (4) the need to represent more than one language using the same script. Highlighting these challenges, some grapholinguistic researchers have pointed out that *akṣara*-based writing systems must be viewed as "fully vowelled *akṣarik* segmentaries" and that the syllable-*akṣara* mapping breaks down in the case of complex *akṣaras* (Gnanadesikan, 2017; 2021)). Others opined that writing-system typologies should consider factors such as linear/non-linear, integral/decomposed, complete/reduced, levels of graphic complexity, and semiotic heterogeneity arising from the use of graphemes from other languages (Fedorova, 2021). Iyengar (2023) pointed out that there is a need to examine the post-consonantal environments in categorizing vowelled segmentaries as abugidas or alphasyllabaries or abugidic-alphasyllabaries. Since the graphematic boundaries of an *akṣara* may not always align with the phonological syllable, one should use graphematic criteria in determining its functional value. In other words, while *akṣara* is associated with phonology, it should not

be decided by it (see Iyengar (2024) for a detailed discussion of this point).

In the past two decades of the 21st century, some grapholinguistic researchers have also argued that there is a need to go beyond representational mapping, which is assumed to be static across all scripts within a given writing-system typology. This should be done by conducting systematic research into the structural properties of graph/grapheme inventories, especially with reference to non-alphabetic writing systems, to specify graphetic and graphematic allography and graphotactic rules. This requires examination of visual complexity to identify which segmentable orthographic units have functional value within a given script, as argued by Meletis (2020a,b). Meletis and Durscheid (2022) commented further that future grapholinguistic research should uncover spatially based systematicity in graph inventories besides giving up on alphabet-specific definition of the construct 'grapheme' because in some writing systems, not all readily differentiated segments function as graphemes. Adopting a multi-modular writing system comprising a graphetic, graphematic (with connections to the language module), and orthographic modules, these researchers proposed the term 'basic shape' to denote a template corresponding to a given graph. Since Abugidas raised a number of questions for defining a grapheme, these researchers offered three distinct criteria for arriving at a universal definition of grapheme:

(1) Lexical distinctiveness (that contributes to meaning differentiation),
(2) Linguistic value (basic shape corresponds to a specific linguistic unit), and
(3) Minimality (basic shape occupies exactly one segmental space).

Chang, Chen, and Perfetti (2018) proposed a multidimensional measure of visual complexity referred to as the 'GraphCom' that measures four different dimensions of visual complexity: (1) Perimetric Complexity (PC), (2) Disconnected Components (DC), Connected Points (CP), and Simple Features (SF). They have empirically demonstrated that a combined measure of graphic complexity enables comparison across writing systems, even those within the same typology. Their database consisted of 131 writing systems and the associated scripts (including the Telugu script). Measures such as the Graphcom will enable researchers to study the interaction between structural features and cognitive processes involved in reading and writing by allowing assessment of the orthographic processing abilities of beginning (bilingual) readers and/or those experiencing reading difficulties. Keeping these applications in mind, we have used the same four dimensions of graphic complexity as Chang, Chen, and Perfetti (ibid.) did to quantify the graph and grapheme complexity of Telugu akṣaras reported in this article.

One of the psycholinguistic variables affecting spoken and/or visual word recognition in alphabetic and non-alphabetic writing systems is the size and density of Orthographic Neighborhood (ON). Marian (2017) reported orthographic and Phonological Neighborhood (PN) databases for many European languages. It was also observed that lexical decision tasks are completed quickly and accurately for words with large neighborhoods (see Davis, 2005; Grainger, 2005; Peereman and Content, 1997). Other researchers reported that large ON will lead to relatively fast eye movements during reading (Yates, Locker, and Simpson, 2004). ON size reportedly affects reading recognition of words even in non-alphabetic scripts associated with languages such as Chinese (Chang, Welbourne, and Lee, 2016; Xiong, Zhang, and Ju, 2021; Zhao, Li, and Bi, 2012). We are not aware of any published research on the ON in relation to any *akṣara*-based Indic scripts, which motivated our study.

Drawing inspiration from some of the grapholinguistic and psycholinguistic research cited in this section, we have undertaken the present study to seek answers to two specific questions relating to the current-day Telugu script:

(1) How does visual complexity distinguish Telugu graphs and graphemes?
(2) What is the relationship between graphemic complexity and ON?

Meletis (2020a) offered definitions of a number of terms relating to both theoretical and applied grapholinguistics. We have used the terms 'graph', 'basic shape', and 'grapheme' in the sense he defined them. The *akṣara* is also referred to as 'orthographic syllable' and enclosed in <-> to distinguish it from the phonological syllable. While we acknowledge the fact that the term 'writing system' encompasses graphetic, graphematic, and sociolinguistic (orthographic) components, our discussion in relation to the Telugu writing system in this paper is restricted for the most part to the graphetic and graphematic components and not to the orthographic module specifying rules of spelling Telugu words correctly. We have adhered to ISO 15919 standards when providing transliteration of the Telugu data included in this paper. The rest of the contents of this paper are organized under the following main headings: Telugu Writing System §2; Graph and Grapheme Complexity §3; Grapheme Complexity and Orthographic Neighborhood size §4, and Implications and Directions for Future Research (§5).

2. The Telugu Writing System

Telugu is one of the four major literary languages belonging to the Dravidian family included in the 8th schedule. It is the official language in two southern states of India, viz., Andhra Pradesh and Telan-

gana. Many tribal languages, such as Gondi, Koya, Konda, and Kolami, use the Telugu script (see https://scriptsource.org for more details). The current-day Telugu writing system may be considered an abugida-alphasyllabary because, first, the main graphic unit '*akṣara*' (CV_0) is an orthographic syllable with an inherent vowel <a>. Secondly, a series of vowel-modifying diacritics (*mātras*) is attached to the basic *akṣara* (CV_0), generating hundreds of orthographic syllables that function as graphemes (see Iyengar (2023) for a detailed discussion of this newly proposed hyphenated typological term). The commonly occurring graphic syllables in native Telugu words are V, Vṃ, CV, CCV, CCVṃ, CCCVṃ. In other words, while all native Telugu words end in open syllables or with a coda-nasal, words borrowed from other languages can have consonant sequences word initially or medially, and a pure consonant without a vowel word finally. Not all *akṣara*s in Telugu script correspond with phonological syllables.

2.1. Inventory of Graphs (Basic Shapes)

The direction of the Telugu writing system is left to right. Words are separated by spaces. All graphic syllables in Telugu are round, each constructed by circles, loops, hooks, and check marks.

In the Telugu script that is currently in use, there are 16 independent vowels, including short and long ones. Vowel length is phonemic in Telugu. Separate graphs are used to represent nasalization and aspiration features. The vowel muting symbol (*virāma* when applied to any consonant, kills the inherent vowel, resulting in a pure consonant that typically occurs in word-final positions in words borrowed from English or Sanskrit. The anuswara <aṃ> and aspiration marker <aha> appear linearly on the horizontal axis. The vowels are represented explicitly in their primary form in word-initial positions. In word-final position, and in borrowed words, they are subordinated to consonants since they are represented by vowel-modifying diacritics. The secondary symbols or diacritics corresponding to all the vowels are referred to as *mātras.* Each one has a specific name that is used during literacy instruction offered during primary school years. When these symbols are applied to basic shapes of <CV_0> akṣaras, the inherent vowel is suppressed, and the akṣara's shape gets altered slightly.

There are 36 graphic symbols to represent basic shapes of CV- *akṣara*s (unaspirated and aspirated) in the Telugu script. In all these *akṣaras*, the inherent vowel /a/ is represented by a check mark, [✓] referred to in Telugu as *talakaṭṭu* 'head-stroke'. In the case of the graph <ట ṭa> it is indicated by a small vertical bar written on top of the left loop; The *akṣara*s corresponding to <జ ja>, <జ్ఞ gna>, <ణ ṇa>, <బ ba> and <ల la> are devoid of overt symbol for the inherent vowel <a>. That is, the *talakaṭṭu* or check

A. Basic shapes

అ	ఆ	ఇ	ఈ	ఉ	ఊ	ఋ	ౠ
a	ā	i	ī	u	ū	ṛ	ṝ
ఎ	ఏ	ఐ	ఒ	ఓ	ఔ	అం	అః
e	ē	ai	o	ō	au	am	ah
క	ఖ	గ	ఘ	ఙ			
ka	kha	ga	gha	ṅa			
చ	ఛ	జ	ఝ	ఞ			
ca	cha	ja	jha	ña			
ట	ఠ	డ	ఢ	ణ			
ṭa	ṭha	ḍa	ḍha	ṇa			
త	థ	ద	ధ	న			
ta	tha	da	dha	na			
ప	ఫ	బ	భ	మ			
pa	pha	ba	bha	ma			
య	ర	ల	వ	ళ			
ya	ra	la	va	ḷa			
శ	ష	స	హ	ఱ			
śa	ṣa	sa	ha	ṟa			

B. Vowel modifiers

◌ా	◌ి	◌ీ	◌ు	◌ూ	◌ృ	◌ె	◌ే	◌ై	◌ొ	◌ో	◌ౌ	◌ం	◌ః
ā	i	ī	u	ū	ṛ	e	ē	ai	o	ō	au	am	ah
కా	కి	కీ	కు	కూ	కృ	కె	కే	కై	కొ	కో	కౌ	కం	కః
kā	ki	kī	ku	kū	kṛ	ke	kē	kai	ko	kō	kau	kam	kah

C. Examples of *akṣaras* with ligatures for geminates and clusters

kka క్క, cca చ్చ, tta త్త, lla ల్ల, wwu వ్వు, rre ర్రె

dru ద్రు, kti క్తి, tsa త్స, tru త్రు, vy వ్య, tri త్రి

Fig. 1. Inventory of graphic elements in the Telugu script

mark is absent in them. In these and all other graphs, the inherent vowel /a/ can be deleted completely, turning them into pure consonants <జ్ j>, <ణ్ ṇ>, <బ్ b> <ల్ l> required for representing words from other languages especially English. The aspirated consonants and fricatives have entered modern-day Telugu through borrowings from Sanskrit, Persian-Arabic, and English. All these *akṣaras* serve as articulatory units during the acquisition of beginning reading.

The graphic forms associated with consonant sequences are referred to as *ottulu* in Telugu. These ligatures apply to *dvitvākṣarālu* (geminates) or *samyuktākṣarālu* (clusters). They typically lie below the basic *akṣara* and are almost never attached to it, unlike the diacritics associated with the vowels. It should be noted that accurate representation in writing requires writing the primary graphs (basic shapes) first and then placing the secondary forms (ligatures) below them. *Akṣaras* made up of geminates tend to occur word medially or word finally, never at the beginning of words. Most words borrowed from Sanskrit have complex onsets with two or even three consonants at the beginning of the word (CCCV). When a vowel is pronounced after a conjunct consonant (as in geminates and clusters), its diacritic is attached to the first and not the second consonant, resulting in some degree of nonlinearity between spoken and written syllables.

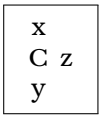
2.2. Visuospatial Logic Governing Telugu *Akṣaras*

In the example *akṣaras* shown below, there is a non-linear relationship between spatial and temporal order conditioned by the placement of the vowel diacritic (shown in red color).

FIG. 2. Telugu *akṣaras* <spi> and <sphū>

The two orthographic syllables with clusters shown above correspond with two CV-phonological syllables, **spi**- and **sphū**- respectively. However, in the right *akṣara* <**sphū**> the orthographic syllable is visually

more complex because the vowel diacritic corresponding to <ū> and the ligature of <sph> occupy larger segmental space compared to the akṣara <spi> on the left. The visuo-spatial logic behind vowel modifiers and ligatures involved in representing geminates and clusters in Telugu was specified by Krishnamurti and Gwynn (1985) as under:



where the primary consonant is represented as C, which has three distinct positions (x, y, and z) around it:

- Position 'x' is exclusively fixed for vowel diacritics attached to the C.
- Position 'y' is fixed for ligatures corresponding to geminates and clusters
- Position 'z' for some secondary symbols of vowels and consonants.

Krishnamurti and Gwynn (1985, p. 34) stated that the positions x, y, z, as well as the degree of contiguity (fused vs. non-fused to the C), distinguish consonant and vowel *akṣaras* besides imposing a strict order in reading them in slow speech in Telugu. They added further that while hundreds of CV syllables and an equal number of Consonant + Consonant combinations are theoretically possible, about 400 different orthographic syllables are sufficient to represent most sounds and sound combinations in Telugu. Referring to the two complex *akṣaras* discussed earlier, viz., <spi> and <sp**hū**>, the spatial location of diacritics for <spi> involves two positions x and y. Whereas, that for <sp**hū**> involves, x (the inherent vowel <a>), y (secondary symbol for <ph>) and z (vowel diacritic for long <u>) attached to the basic akṣara <s> on the horizontal axis, contributing to higher level of visual complexity of akṣara <sp**hū**> compared to that of <spi>. It should be noted that this logic works for a vast majority of *akṣaras* although there are some exceptions relating to fusion or non-fusion of *akṣara*'s components with the base as in these two words as well as words such as స్నేహం snēham 'friendship' in which the vowel modifiers in both the *akṣaras* are not attached to the base consonants, <స sa> and <హ ha> respectively in this case. Whether graph complexity varies depending on whether the vowel-modifying diacritics are fused with the base (CV_0) is one of the points we wish to probe in this study.

2.3. Graphemes in Telugu Script

This section provides information on how the basic shapes of individual graphs in the script inventory combine to form graphemes that make up meaningful Telugu words.

Monosyllabic meaningful words are extremely limited in Telugu: Some examples include:

rā	రా	'come'	*ā*	ఆ	'that'
pō	పో	'go'	*ī*	ఈ	'this
nā	నా	'mine'	*ē(ṃ)*	ఏం	'why'
mī	మీ	'yours'	*strī*	స్త్రీ	'woman'

It should be noted that in the first five words, the diacritic associated with long vowels is attached to the base CV-*akṣara*, whereas in the bottom two words, it is not. Since they do satisfy the criteria of mapping onto one phonological syllable, they are lexically distinctive and occupy one segmental space; we can treat them as 'graphemes.'

Most *akṣara*-based writing systems, including Devanagari associated with Hindi, have independent vowels appearing in word-initial position, and dependent vowels (diacritics) elsewhere. The independent vowels and their post-consonantal diacritics are to be treated as allographs since they are in complementary distribution and correspond with the same linguistic unit. The initial phase of literacy instruction in Telugu is always aimed at teaching children to practice pronouncing and writing the basic shapes of CV *akṣaras* listed in the *varnamāla* by applying various vowel-modifying diacritics or *guṇintālu* as well as ligatures or *ottulu* even if they lack meaning. In fact, specially designed copyright booklets are given to children to fill in the basic shapes appearing with vowel diacritics and ligatures.

The basic shapes of akṣaras combine with each other to serve graphematic function in two-syllable meaningful words, as the examples listed below illustrate:

- <క ka> combines with <కి ki> to become a meaningful word, కాకి *kāki* 'crow'.
- <కం kaṃ> combines with <ద da> to become the word, కంద *kanda* 'yam'.
- <కు ku> combines with <క్క kka> to become కుక్క *kukka* 'dog'.
- <కు ku> combines with <ట్ర ṭra> to become కుట్ర *kuṭra* 'conspiracy'.
- <క్రు kru> cobines with <షి ṣi> to become కృషి *kruṣi* 'effort'.

These examples illustrate that all basic shapes with vowel diacritics and those with ligatures can serve as graphemes in meaningful words, fulfilling the criterion of lexical distinctiveness. Each *akṣara* also corresponds to one phonological segment, and most of the *akṣara*s occupy one segmental space, with one exception, that of homorganic nasals indicated by a zero which lie next to the consonant on the horizontal plane (never attached to it) as in the word, కంద *kanda* 'yam' which aligns with the phonological syllables /kan/and /da/ or అందం andam 'beauty' with phonological syllables /అం am/ /దం dam/. It should be noted that in words with consonants appearing in coda position, a timing unit called

mora plays a role in processing and segmentation tasks. Mora is larger than a phoneme but smaller than a syllable. Patel (2007, p. 173) pointed out that the distinction between open and closed syllables and the concept of syllable quantity is implicit in the design of written *akṣaras*.

We reproduce below an excerpt from a Telugu newspaper to show different types of *akṣaras* that serve as graphematic syllables/words in the current-day Telugu script (each *akṣara* is separated by a period mark and each word by a hashtag #):

(1) న. వ. తె. లం. గా. న. లో # స. ర్కా. రీ # కో. లు. వు. ను. చే. జి. క్కిం.
na. va. te. lam. gā. na. lō # sa. rkā. rī. # ko. lu. vu. nu. cē. ji. kkin.
చు. కో. వ. డ. మే # ల. క్ష్యం. గా # క. ష్ట. ప. డి చ. దు. వు. తు.
cu. kō. va. ḍa. mē # la. kṣyam. gā # ka. ṣṭa. pa. ḍi ca. du. vu. tu.
న్న. అ. భ్య. ర్థు. లు # ఎ. ప్పు. డె. ప్పు. డా
nna. a. bhya. rdhu. lu # e. ppu. ḍe. ppu. ḍā

నవతెలంగానలో సర్కారీ కొలువును చేజిక్కించుకోవడమే లక్ష్యంగా కష్టపడి చదువుతున్న అభ్యర్థులు ఎప్పుడెప్పుడా

The aspirants studying hard with the primary goal of getting a government job in the newly formed state of Telangana are waiting

Among the three words in red, viz., *sarkāri* 'government' in the first line is a Hindi word, whereas *lakshyaṃ* 'goal', in the first line, and *abhyardhulu* 'aspirants' in the second line respectively are words borrowed from Sanskrit. These words have been assimilated into Telugu with certain modifications to the graphs to accommodate morphological-level information, such as plurality and case. When they are written in Devanagari script associated with Hindi, they appear as shown: सरकारी, लक्ष्य and अभ्यर्थी in direct correspondence with their phonological forms. The arrangement of graphemes in these two Indic scripts differs, suggesting differences in representational mapping. To elaborate, using the Hindi word *sarkāri*, while the Devanagari script represents <K> in its primary form, the Telugu script uses the secondary form of <k> in the second syllable. The Sanskrit word *lakṣyaṃ* is also modified from the original version in Sanskrit by replacing <a> with <aṃ> and adding the adverbial suffix *-gaa* in Telugu. The Sanskrit word, अभ्यर्थी, when written in Telugu, also gets modified by the addition of a plural marker *-lu* required in the context of this sentence. Thus, the arrangements of graphs belonging to languages other than Telugu appear to follow Telugu graphematic rules, sometimes even by violating phonotactic rules. The graphematic rules in the case of complex graphemes in Telugu are driven by spatial logic more than phonological information, a fact offering support to the suggestion that *akṣara* should be treated as a graphematic unit by Iyengar (2024).

In the sample of Telugu writing printed above, there are 43 different orthographic syllables (*akṣaras*) altogether. Of these, 30 have the prototypical V, CV(V), CVṃ structures, while the others containing ligatures representing geminates / clusters have C_1C_1V or C_1C_2V(m)-type structures. Looking at the example, ల. క్ష్యం lakṣyaṃ 'goal', it has the structure CV. $C_1C_2C_3$ *ṃ*. The first akṣara <ల la> having a CV structure aligns with one phonological syllable. When it is replaced by <భ bha> another meaningful word, *bhakṣyaṃ* 'a sweetmeat' obtains, and thus it is involved in lexical distinctiveness. It occupies one segmental space. Since it satisfies all three criteria specified for 'grapheme' by Meletis (2020a), it can be treated as a simple CV grapheme.

The second *akṣara*, <kṣyaṃ>, on the other hand, demands a closer examination of the arrangement of basic shapes. The complex akṣara <*kṣa*> has a graphematic function by itself in that it appears in many meaningful words such as: *lakṣa* 'hundred thousand', *pakṣi* 'bird', *śikṣa* 'punishment', *mōkṣaṃ* 'liberation', *rakṣaṇa* 'protection' *kṣaya* 'Tuberculosis' and so on. The primary form of <k> and the secondary form of <ṣ> are stacked vertically, and yet they occupy one segmental space, qualifying all the criteria to be considered as a complex grapheme. When it combines with the next akṣara, <yam>, it becomes a bound grapheme. The word lakṣyaṃ can also be written as: లక్ష్యము lakṣya*ṃu* or లక్ష్యమ్ with three syllables. It is instructive to cite here the comments of Iyengar (2024, p. 431): "If *akṣara* is conceived as a written unit centered around one free graph, it follows that one can add or remove as many bound elements as graphematically permissible without impacting the *akṣara* count. More research is needed into the *akṣara's* compositional transparency in Telugu script."

2.4. Graphematic Features in Telugu Script

With a view to examining some graphematic relations governing the akṣaras in the Telugu script, we have selected 200 two-syllabic words from a text corpus. This corpus of over 3 million words of running text, covering modern fiction, short stories, novels, science writing, children's stories, and journalese, was developed by the Central Institute of Indian Languages, Mysore, India (CIIL corpus, henceforth). A detailed analysis of the 1000 most frequently appearing printed words, orthographic syllables, and character sequences, along with their frequency counts in relation to the CIIL corpus, was reported in Uma Maheshwar Rao (2006). We have chosen 200 out of this list of 1000 words for our analysis of the visual complexity of *akṣaras*. In conducting this analysis, we drew on the spatial logic governing Telugu *akṣaras*, as discussed earlier in this paper, as articulated by Krishnamurti and Gwynn (1985). Specifically, the basic *akṣara*, <CV_0> has most of the vowel modifiers attached to it on top

(position X). Most ligatures representing consonant sequences (geminates and clusters) lie below the basic akṣara and are rarely attached to it (position Y). Homorganic nasals, aspiration marker (visarga) and vowel modifier corresponding to the grapheme <ఉ u / ఊ ū> lie on the horizontal plane next to <CV_0> at position Z. If we disregard the checkmark corresponding to the inherent vowel that is present in most *akṣaras* and classify them based on the position of other vowel modifiers and ligatures, ten different orthographic syllable types cover most of the 200 target words. The number of words belonging to a given akṣara type is indicated under N= in the first column in Table 1 below:

The following facts are evident from Table 1:

1. There are 25 words of the type V. CV / V. CVV such as ఒక *oka* 'one' and ఎలా *elā* 'how'. Such *akṣaras* are fully aligned with the phonological syllables and may be treated as simple akṣaras.
2. Words with homorganic nasal containing vowel modifier belonging to /u/ as in the word, మందు *mandu* 'medicine' (3rd word from top of the Table 1) corresponds directly to the phonological syllable *man.du.*
3. In contrast to types (1) and (2), the last row of Table 1 has only 9 words with *akṣaras* of the type, C_1C_2V. C_1C_2V (*ṃ)* (e.g., వ్యక్తి <*vya.kti*> 'person'). This word is phonologically syllabified as /vyak/. /ti/ in compliance with the phonotactic rules of Telugu. However, it does not correspond with the orthographic syllable division.

One other fact that emerges from this analysis is that when the secondary graphs representing geminates or clusters that appear in position Y (bottom of the basic *akṣara*), the alignment between orthographic and phonological syllables breaks down as in words, అమ్మ *amma* 'mother', పెద్ద *pedda* 'big',చుట్ట *cutta* 'cigar' *and* and those with clusters: రక్తం *raktam* 'blood', దృష్టి *druṣṭi* 'vision' and స్పష్టం *spaṣṭaṃ* 'clarity' (listed in rows 5-10 of Table 1) suggesting that presence of ligatures seems to contribute to increase in visual complexity of Telugu *akṣaras.* This analysis, based on 200 two-syllable words, suggests that at least three major types of graphemes operate in Telugu script:

(1) Simple (V(V), Vṃ, CV, CVṃ, CVCV, CVṃ.CV)
(2) Complex (V.CCV, CV.CCV) and
(3) Compound (CV. C_1C_2V; C_1C_2Vṃ. C_1C_2V; and C_1C_2V. C_1C_2Vṃ)

The third type has many bound graphemes. The graphotactic rules for the two-syllabic native Telugu words permit the following types of *akṣaras*:

(1) Simple syllable alone as in for example, రా *raa* 'come'
(2) simple syllable + ligature corresponding to a geminate as in అమ్మ *amma* 'mother' or అద్దం addaṃ 'mirror'

Table 1. Arrangement of graphematic units within *akṣaras*

Orthographic syllable structure	Position of the V-diacritic and ligatures	Telugu Word with Transliteration	English Gloss	Status of Spatiotemporal alignment
V.CV(V) N=25	X in second syllable	idi ఇది	'This one'	Present
Vṃ.CV N=14	Z in the second syllable	andu అందు	'In it'	Present
CVṃ.CV N=25	Z in both syllables	mandu మందు	'Medicine'	Present
CV(V).CV N=50	X in the 2nd syllable	pani పని	'Work'	Present
V.C_1C_1V N=08	Y in the second syllable	amma అమ్మ	'Mother'	Absent
CV.C_1C_1V N=15	X first syllable, Y in second syllable	pedda పెద్ద	'Big'	Absent
CV.C_1C_1V N=10	Z in first syllable, Y in second syllable	cuṭṭa చుట్ట	'Cigar'	Absent
CV.C_1C_2Vṃ N=25	Z and Y in the second syllable	raktaṃ రక్తం	'Blood'	Absent
C_1C_2V(ṃ). C_1C_2V N=16	Y in first syllable, X and Y second syllable	druṣṭi దృష్టి	'Vision'	Absent
C_1C_2V.Vṃ N=09	Y in first syllable, Y and Z in second syllable	spaṣṭaṃ స్పష్టం	'Clarity'	Absent'

(3) simple syllable + complex / compound syllables as in లక్ష lakṣa 'hundred thousand' *or* లక్ష్యం lakṣyaṃ 'goal'.

Fedorova (2013) also made a distinction among simple, complex, and compound *akṣaras* and stated that the degree of complexity within these three types of structures varies across different abugida scripts, a point that needs to be probed in future research on Indic scripts. In a more recently published paper, Fedorova (2021) commented that vowels in an abugida act like different "garments" for consonants, sometimes they form the "soul" of a consonant "body" in an *akṣara*. We feel that this observation aptly captures the *akṣara*, kṣyaṃ in Telugu script because the heavy vowel <*aṃ*> applies simultaneously to the three consonants: <k>, <ṣ>, and <y>.

Words borrowed from Sanskrit such as: దృష్టి *druṣṭi* 'vision' and స్పష్టం *spaṣṭaṃ* 'clarity' with syllable structure C_1C_2V. C_1C_2V in both *akṣaras* are also to be treated as compound graphematic words with the greatest amount of graph/grapheme complexity compared to words with relatively simple orthographic syllable structure involving individual basic shapes, such as for example, ఋషి ṛuṣi 'sage' or patam పటం 'frame' or శాపం śāpam 'curse' without ligatures. The visuospatial analysis of Telugu *akṣaras* discussed in this section suggests that modifications to the base form <CV_0> due to the presence of ligatures enhance the visual complexity of graphemes. This observation is substantiated by the quantitative analysis discussed in the next section. Characteristic features of the current day Telugu Script are listed in Appendix A.

3. Graph and Grapheme Complexity

This section provides quantitative information regarding graph and grapheme complexity (GC) as measured in our study.

3.1. Graph Complexity

As the programming code for estimating GC measures reported by Chang, Chen, and Perfetti (2018) is not available, nor are the actual estimated values for Telugu akṣaras reported by them, we reimplemented both in Python (available at the GitHub repository: `https://tinyurl.com/37y9ae2a`). We have used the same four graphic complexity measures proposed by Chang, Chen, and Perfetti (ibid.), namely, the number of Connected Points (CP), the number of Simple features (SF), the number of Disconnected Components (DC), and Perimetric Complexity. CP is defined as the number of intersection points between SFs. SF refers to

the number of loops, lines, dots, and curves. DC is taken as the number of disconnected points. Finally, PC is defined as the ratio of the squared perimeter of a graph (number of pixels) to the number of background pixels in the graph. Specifically, PC is $\frac{P^2}{A4\pi'}$, where P is the sum of the inside and outside perimeters of the foreground (informally, it is the sum of the perimeters of the inked sub-regions of the akṣaras), A is the foreground area (Pelli, Burns, Farel, and Moore-Page, 2006; Watson, 2012). We illustrate how these measurements are performed using a single graph, <ఘ> 'gha', shown below in Figure 3.

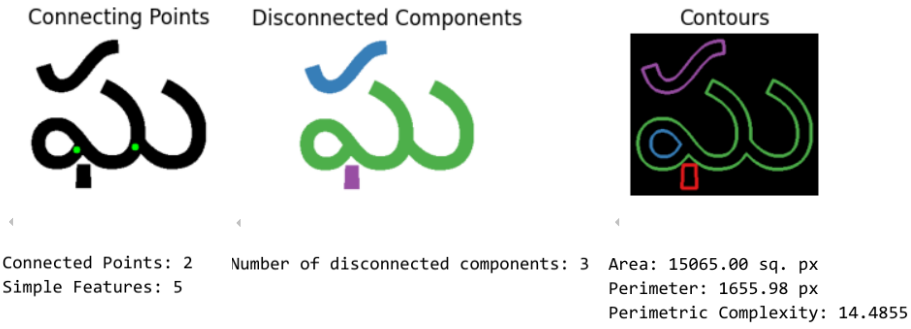


FIG. 3. Graphic complexity measures applied to the graph <gha>

The actual values for all basic shapes in the Telugu script are included in Appendix B. These four graphs show that for all the basic shapes of Telugu akṣaras,

- Simple Features (SF) range from 2-8;
- Connected Points (CP) range from 0-7;
- Disconnected Components (DC) from 1-3; and
- Perimetric Complexity (PC) ranges from 7.66 to 22.32.

Looking closely, one can discern a pattern according to which:

1. the visual complexity is the highest among vowels <ṛu>, <*ṝū*>, <am>, <aha>, <ū> and diphthong <au> followed by
2. aspirated consonants and then unaspirated consonants. It may be reasonable to argue that words containing aspirated consonants are visually more complex than those with unaspirated components because they contain more disconnected components due to the presence of the aspiration marker indicated below the base.
3. Graphs in which vowel modifiers are not attached to the top of the base CV_0, appear to possess greater graph complexity over those in which the vowel modifiers are attached.

We have validated our result for one graph with those reported by Chang, Chen, and Perfetti (2018) for the *akṣara*, <aha>, see Table 2.

Table 2. Graphic complexity values for <aha>

GC dimension	aha అః Present study	aha అః Chang et al. (2018) study
PC	16.70*	18.06*
DC	3	3
CP	2	2
SF	5	5

The perimetric complexity is the greatest for *akṣara* <aha> compared to all the other vowels, a finding in agreement with that reported by Chang, Chen, and Perfetti (ibid.) as shown in Table 2. It should be noted that the slight difference in PC values between the two studies is due to differences in the font types used. We have used the Noto Sans Telugu Regular font, whereas Chang, Chen, and Perfetti (ibid.) used the NATS font.

The *akṣaras* that do not have a checkmark on top have fewer connected points as well as fewer disconnected components (for example: <ల la> and బ ba>). On the other hand, those that have a connected checkmark have fewer DCs but more CPs (for example, <గ ga, క ka, ర ra>) and vice versa for akṣaras with disconnected checkmarks (for example <ఏ ee>, <స sa>, <ప pa>).

The next table (Table 3) shows the changes in graphic complexity as the grapheme <క ka> gets modified when vowel modifiers and ligatures are added. As can be seen, all the GC measures tend to increase when a basic shape is transformed by modifiers.

Table 3. Graphic complexity values for <ka> with vowel diacritics and ligatures

Complexity Dimension	క <ka>	కో <kō>	క్క <kka>	క్ల <kla>
PC	8.74	14.19	16.69	14.18
DC	1	1	2	2
CP	1	3	2	2
SF	2	5	4	4

Notice that the PC value increases to a greater extent when <క ka> appears as a double consonant <క్క kka> compared to the cluster, <క్ల kla>.

It could be related to the fact that the secondary form of <k> when it is doubled takes up more space, occupying position 'y' below the base consonant, and position 'z' on the horizontal axis. The secondary form (ligature) of <ల la>, however, remains below, but quite close to the base consonant graph. It is also relatively small, compared to that associated with the graph representing <క్క kka>. In actual writing, the exact position of the ligatures varies considerably. For example, the word, *druṣṭi* 'vision' can be written either as: దృష్టి or as ద్రుష్టి.

3.2. Grapheme Complexity

When the basic *akṣaras* join to form meaningful 2-syllable words, we have Grapheme Complexity. Grapheme Complexity is estimated by considering the image of two orthographic syllables written together and then estimating various GCs as defined in the previous section. As shown in Table 4. As grapheme complexity increases, GC values (particularly PC) also increase, as evidenced by three of the 200 target words listed in Table 4. However, it should be noted that DC, CP, and SF do not strictly follow this linear trend in several cases.

TABLE 4. Grapheme complexity values for three Telugu words with different syllable structure

GC Dimension	adi V.CV అది 'that'	mūḍu CV.CV మూడు 'three'	vyakti C_1C_2V.CV వ్యక్తి 'person'
PC	20.66	33.97	35.12
DC	2	2	4
CP	5	12	5
SF	6	14	11

It should be noted that Telugu words with simple *akṣaras* (V.CV) have the least PC values compared to those with graphematic syllables <**mū**> and <**ḍu**> (CVV.CV) and those with even more complexity (C_1C_2V.C_1C_2V). It appears that the position of the vowel modifier also affects graph complexity. The next step is to examine the relationship between grapheme complexity and the size of orthographic neighbors for each target word.

4. Grapheme Complexity and Orthographic Neighborhood Size

A Telugu word's neighbors are those which differ from the target word in one entire *akṣara* or any of the sub-components of the *akṣara*, as the examples below illustrate:

- Substitution of an entire *akṣara* in the initial position of the word, గుల gula 'itch,' results in a new meaningful word, వెల vela 'price. This is an example of a phonological neighbor (PN).
- Deletion of the homorganic nasal from the word, తొండ *toṇḍa* 'chameleon' results in తొడ *toḍa* 'thigh' is an example of ON.
- Addition of a segment (ligature corresponding to <గ ga> to the word, వరం *varaṃ* 'boon' can result in వర్గం *vargaṃ* 'group,' another example of ON.

This task was done using a specially designed algorithm applied to an online corpus. There were four main steps in the algorithm: (1) Analyze the input word's orthographic syllables, (2) Process each syllable, (3) Find matching words, and (4) Remove duplicate words and return the results. The lexical status of all the neighbors generated through these operations is verified using an online dictionary (`https://dsal.uchicago.edu/dictionaries.brown`). It should be noted that some dictionary entries belong to both old Telugu and Sanskrit.

With the help of this algorithm, all the 200 two-syllable words were subjected to substitution, deletion, and addition operations such that the result is yet another meaningful word that exists in the dictionary. It should be noted that if we had included words longer than two syllables, we could have obtained more information on orthographic neighbors, since deletion works better with three- or four-syllable words. Based on the data we have analyzed, we cannot specify the relationship among frequency, word length, and ON size, a topic that warrants further research.

A systematic comparison was made of all 200 words and their corresponding ONs. We have generated word clouds for words with simple, complex, and compound graphemes.

- Words such as ఆట aaṭa 'play' had 154 neighbors.
- Words such as రెక్క rekka 'wing' with a ligature had 93 ONs.
- Words such as పూర్వం puurvam 'past' had only 31 neighbors.

These results clearly show that as the visual complexity of *akṣara* increases, involving vowel-modifying diacritics or ligatures corresponding to consonant sequences, the grapheme's complexity increases. Some limitations of our study should be noted. We have used only 200 high-frequency two-syllable words and conducted all the analysis using only the CIIL corpus. However, the results need to be verified for the other

two-syllable words and extended for multi-syllable words. The code in the repository is useful for accurately estimating DC and PC, but needs further refinement to estimate SF and CP. The syllable structure extraction and generating ONs are scalable to multi-syllable settings. We faced considerable difficulty in finding a comprehensive, free-access text corpus and a contemporary online Telugu Dictionary with a lookup facility for the algorithm to verify whether the generated ON is a valid word.

Statistical analysis of our data revealed a negative relationship between grapheme complexity and ON size, as evidenced by the numbers, especially those corresponding to DC and PC, with the strongest negative correlation with the ON shown below. The scatter plot in Figure 4 below exhibits this relationship between PC and ONs:

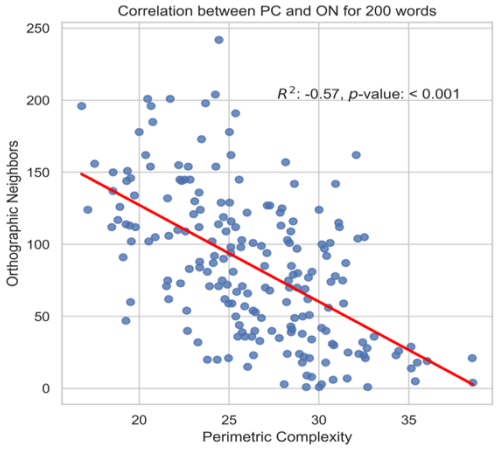


Fig. 4. Negative correlation between Perimetric Complexity and ON

The regression plot in Figure 4 shows a negative correlation between perimetric complexity and orthographic neighborhood size, such that the greater the graphemic complexity of orthographic syllables (akṣaras), the smaller the size of ON for Telugu.

Researchers have argued that grapheme complexity is strongly and positively associated with learning difficulties, and that the extent of this difficulty varies depending on whether the writing system is associated with the native language or a later-learned second language. Perfetti and Verhoeven (2022) argued that the visual complexity of graphs / graphemes is a prime candidate for a non-mapping factor that could make a difference to reading ability.

5. Implications and Directions for Future Research

We conducted this study to address two main questions: (1) How does visual complexity distinguish Telugu graphs and graphemes? And (2) What is the relationship between graphemic complexity and ON? Two of the four dimensions of graphic complexity we measured, viz., Perimetric Complexity (PC) and Disconnected Components (DC), demonstrated variation in both graphic and grapheme complexity across the entire inventory of akṣaras in the Telugu script. We have also shown that, for two-syllable words in Telugu, the higher the grapheme complexity, the smaller the size of the orthographic neighbors containing those akṣaras. Both these findings have many implications, both theoretical and clinical.

As mentioned earlier, the literacy learning environment in multilingual India is vastly different from most other countries in the Northern hemisphere. Depending on their socioeconomic status and parental education, children have varying degrees of exposure to oral languages and access to print materials at home. Yet, when they enter school, they are required to learn almost simultaneously one or two akṣara-based Indic scripts, along with English, which is associated with an alphabet. While there is a consensus among second language writing system researchers that specific combination of writing systems (and their scripts) has differential effect on literacy acquisition in general, and development of metalinguistic knowledge in particular (see Cook and Bassetti (2010), there is scant published works on how exactly children develop knowledge about the relationship between spoken units of a language and the graphic symbols especially with reference to Indic scripts. Empirical evidence regarding children's orthographic processing abilities across multiple scripts is virtually nonexistent. Most assessment tools are either adapted versions of English-based tests or rely on researchers' insights into the graphic complexity of akṣaras. In this context, there is a danger of misdiagnosing developmental dyslexia among Indian children. The results of our study, when bolstered with further studies involving larger text corpora and longer target words for estimating phonological and orthographic neighbors (density and frequency), will have tremendous implications for gaining better insights about language-script learning, language education practices, and language disorders.

Recent research on writing system diversity involving groups with different cognitive abilities (see, for e.g., Harris and Hirshorn (2022) revealed considerable individual differences in reading abilities at both behavioral and neural levels. The earlier finding that the Visual Word Form Area (VWFA) in the left hemisphere plays a crucial role in skilled word reading is now reported to be restricted to readers of alphabetic writing systems. Readers of non-alphabetic writing systems tend to exhibit bilateral engagement of VWFAs in both hemispheres, perhaps due

to the system's demand for more complex orthographic units characterized by greater visual complexity and stronger connections to semantics rather than only phonology. Since the development of reading strategies depends on specific writing systems, instructional material must be language-specific. There is an urgent need to determine the precise nature of the internal complexity of orthographic representations stored in the lexicon and the orthographic processing abilities of different populations to enhance our understanding of how bi-literacy acquisition works. The graph and grapheme complexity values we report for Telugu can provide much-needed guidance for designing graded reading tests for children with developmental dyslexia and adults with acquired dysgraphia. Some topics awaiting future research aimed at understanding the cognitive consequences of literacy in multiple languages/scripts in the Indian context include:

- Development of tests such as *akṣara discrimination*, word recognition, lexical decision that will allow researchers to manipulate subcomponents of Telugu akṣaras, such as vowel modifiers, homorganic nasals, ligatures, etc., that will facilitate measurement of ak*ṣara* processing abilities.
- Designing behavioral and eye-tracking studies of word recognition when target words are presented in sparse vs. dense neighborhoods for words that appear in school textbooks.
- Future grapholinguistic research in relation to abugidic alphasyllabaries will also benefit from data relating to graphemic frequencies and graphotactic patterns in various Indic scripts.
- Systematic studies that permit comparison among scripts associated with languages that have similar graph/grapheme inventories, such as Telugu vs. Kannada or Tamil vs. Malayalam.
- Development of computational models of *akṣara* naming / discrimination
- Analysis of word reading speeds/accuracies among children with congenital hearing loss, developmental dyslexia, acquired dyslexia, and dysgraphia following brain damage
- Application of graph/grapheme complexity measures to analyze writing impairments in individuals diagnosed with Parkinson's Disease.

Acknowledgments

We thank Arvind Iyengar for sharing with us some of the published papers on this topic and offering useful comments on an earlier version of this paper.

Appendix A

Characteristic Features of Current-Day Telugu Script (as per the Codes Specified in ScriptSource.org)

1. Language	: Telugu [tel], Family: Dravidian
2. Script	: Telugu [Telu]
3. Id.	: te-Telu-IN (Telugu script used in India)
4. Direction	: Left to Right
5. Type	: Fully vowelled akṣarik segmentary
6. Basic akṣaras	: <CV_0> # 35
7. Inherent vowel	: <a>
8. Independent vowels	: Yes, in word initial position
9. Dependent vowels	: In post consonantal position after <CV_0>
10. # of vowel *akṣaras*	: 16
11. Vowel modifiers	: Mostly fused with <CV_0> on top
12. Dominant graph In conjunct *akṣaras*	: C1 to which vowel modifier is attached
13. Ligature for geminates	: non-fused with <CV_0> at the bottom / stacked
14. Ligature for clusters	: non-fused with <CV_0> at the bottom / stacked
15. Obligatory virama (vowel killing sign)	: Yes, mostly used with borrowed words
16. Nasality	: Full form for <m>, <n>, zero post consonantal
17. Visarga	: Yes, mostly with Sanskrit words
18. Non-sanskrit *akṣaras*	: Yes <fa> written as <pha> and <za> as <Ja>
19. Aspiration sign	: Vertical line below <CV_0>
20. Word spacing	: Yes
21. Syllable marking	: No
22. Geminate sign	: No
23. Top mark on *akṣaras*	: yes, check mark on some, absent in others
24. Complex and compound Graphemes	: Yes
25. Grapheme-phonology Mapping	: Syllable and mora

Appendix B

Distribution of Graphic Complexity Values for All the Basic Akṣaras in the Telugu Script

Distribution of Simple Features (SF)

SF values	*akṣaras*
2	ఎ, ఒ, క, గ, న, బ, ర, ల, స
3	అ, ఏ, ఐ, ఓ, జ, ఠ, ద, ప, వ
4	ఆ, ఇ, ఋ, అం, ఙ, చ, ట, డ, ణ, ధ, ఫ, భ, మ, య, ఱ, ళ, శ, ష
5	ఈ, ఉ, ఔ, ఁ, ఖ, ఘ, ఛ, ఝ, ఢ, త, థ, హ
6	ఌ, ఞ
8	ఊ

Distribution of Connected Points (CP)

CP values	akṣaras
0	స
1	ఎ, ఏ, ఒ, క, గ, ఠ, న, ప, ఫ, బ, ర, ల
2	అ, ఆ, ఇ, ఐ, ఓ, అం, ఁ, ఖ, ఘ, ఙ, ఞ, ఢ, ద, ధ, భ, వ, శ, ష
3	ఋ, ఋ, ఙ, చ, ఛ, ఝ, ట, డ, ఢ, ణ, మ, య, ఱ, ళ, హ
4	ఈ, ఔ, ఞ, త
5	ఉ, ఌ
7	ఊ

Distribution of Number of Disconnected Components (DC)

DC values	*akṣaras*
1	అ, ఆ, ఇ, ఈ, ఉ, ఊ, ఋ, ఌ, ఎ, ఐ, ఒ, ఓ, ఔ, క, గ, ఙ, చ, జ, ఞ, ట, డ, ణ, త, ద, న, బ, మ, య, ర, ఱ, ల, ళ, వ, శ
2	ఏ, అం, ఋ, ఖ, ఛ, ఝ, ఠ, ఢ, ధ, ప, భ, ష, స, హ
3	ఁ, ఘ, థ, ఫ

Distribution of Perimetric complexity (PC) values

PC range	*akṣaras*
7-9	ఎ, ఒ, ర, ల
9-11	ఆ, ఇ, ఏ, ఐ, ఓ, క, ఖ, గ, ఙ, చ, ఞ, ట, ఠ, ణ, త, ద, న, ప, బ, వ, శ, స
11-13	అ, ఉ, ఛ, ఞ, డ, ఢ, థ, ధ, ఫ, భ, ఱ, ళ, ష

13-15	బొ, ము, హా
15-17	ఈ, ష, ఘు
17-19	ఊ, ఋు, ఱు, యు
19-21	అం
21-23	ఋూ

References

Bhuvaneshwari, B. and P. Padakannaya (2014). "Reading in Tamil: A more alphabetic and less syllabic akshara-based orthography". In: *South and Southeast Asian Psycholinguistics*. Ed. by H. Winsket and P. Padakannaya. Cambridge: Cambridge University Press, pp. 192–201.

Chang, Y. N., S. Welbourne, and C. Y. Lee (2016). "Exploring orthographic neighborhood size effects in a computational model of Chinese character naming". In: *Cognitive Psychology* 91, pp. 1–23. DOI: 10.1016/j.cogpsych.2016.09.001.

Chang, Li-Yun, Yen-Chi Chen, and Charles Perfetti (2018). "GraphCom: A multidimensional measure of graphic complexity applied to 131 written languages". In: *Behavioral Research Methods* 50, pp. 427–449.

Chatterjee-Singh, Nandini and T. A. Sumathi (2019). "The role of phonological processing and oral language in the acquisition of reading skills in Devanagari". In: *Handbook of Literacy in Akshara Orthography*. Ed. by R. Malatesha Joshi and Catherine McBride. Springer Nature, pp. 261–278.

Cook, Vivian and Benedetta Bassetti (2010). *Second Language Writing Systems*. New Delhi: Viva Books.

Davis, C. J. (2005). "N-watch: A program for deriving neighborhood size and other psycholinguistic statistics". In: *Behavioral Research Methods* 37.1, pp. 65–70.

Duggirala, Vasanta (2019). "Akshara processing in Telugu depends on syllabic and phonemic sensitivity". In: *Handbook of Literacy in Akshara Orthography*. Ed. by R. Malatesha Joshi and Catherine McBride. Springer Nature, pp. 119–138.

Fedorova, Ludmila L. (2013). "The development of graphic representation in abugida writing: The akshara's grammar". In: *Lingua Posnaniensis*, pp. 49–66.

———— (2021). "On the typology of writing systems". In: *Grapholinguistics in the 21st Century. 2020 Proceedings*. Ed. by Y. Haralambous. Brest: Fluxus Editions, pp. 805–824.

Gnanadesikan, Amalia E. (2017). "Towards a typology of phonemic scripts". In: *Writing System Research* 9.1, pp. 14–35.

———— (2021). "Brahmi's children: Variation and stability in a script family". In: *Written Language and Literacy* 24.2, pp. 303–335.
Grainger, J. (2005). "Effects of phonological neighborhood and orthographic neighborhood density interact in visual word recognition". In: *The Quarterly Journal of Experimental Psychology A* 58.6, pp. 981–998.
Harris, Lindsay N. and Elizabeth Hirshorn (2022). *Culture is not destiny for reading: Highlighting variable routes to literacy within writing systems*. Tech. rep. 3. CISLL. URL: https://huskiecommons.lib.niu.edu/ctrcisle-publication/3.
Iyengar, Aravind V. (2023). "More matters of typology: Alphasyllabaries, abugidas and related vowelled segmentaries". In: *Written Language and Literacy* 26.1, pp. 30–56.
———— (2024). "The akshara as a graphematic unit". In: *Grapholinguistics in the 21st Century Proceedings. Grapholinguistics and its Applications*. Ed. by Y. Haralambous. Vol. 10. Brest: Fluxus Editions, pp. 419–436.
Krishnamurti, Bh. and J. P. L. Gwynn (1985). *A Modern Grammar of Telugu*. New Delhi: Oxford University Press.
Marian, Viorica (2017). "Orthographic and phonological neighborhood databases across multiple languages". In: *Written Language and Literacy* 20.1, pp. 7–27.
Meletis, Dimitrios (2020a). *The Nature of Writing: A Theory of Grapholinguistics*. Vol. 3. Grapholinguistics and its Applications. Brest: Fluxus Editions.
———— (2020b). "Types of allography". In: *Open Linguistics*. DOI: 10.1515/opli-2020-0006.
Meletis, Dimitrios and C. Durscheid (2022). *Writing Systems and their Use: An Overview of Grapholinguistics*. Berlin: De Gruyter Mouton.
Nag, Sonali (2014). "Akshara-phonology mappings: The common yet uncommon case of the consonant cluster". In: *Writing Systems Research* 6, pp. 105–119.
———— (2017). "Learning to read languages of South Asia". In: *Learning to Read Across Languages and Writing Systems*. Ed. by Ludo Verhoeven and Charles Perfetti. Cambridge: Cambridge University Press, pp. 104–126.
Nambiar, Krithika, Kiran Kishore, and Pranesh Bhargava (2023). "The effect of script reform on level of orthographic knowledge: Evidence from alphasyllabary Malayalam". In: *PLoS ONE* 18.8. DOI: 10.1371/journal.pone.0285781.
Padakannaya, Prakash et al. (2016). "Visual orthographic complexity of aksharas and eye movements in reading: A study in Kannada alphasyllabary". In: *Writing Systems Research* 8.1, pp. 32–43.
Patel, Purushottam G. (2007). "Akshara as a linguistic unit in Brahmi scripts". In: *The Indic Scripts: Paleographic and Linguistic Perspectives*. Ed. by P. G. Patel, P. Pandey, and D. Rajgor. New Delhi: D.K. Printworld (P) Ltd., pp. 167–213.

Peereman, Ronald and Alain Content (1997). "Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space". In: *Journal of Memory and Language* 37, pp. 382–410.

Pelli, D. G. et al. (2006). "Feature detection and letter identification". In: *Vision Research* 46, pp. 4646–4674.

Perfetti, Charles and Ludo Verhoeven (2022). "Writing systems and global literacy development". In: *Global Variation in Literacy Development*. Ed. by L. Verhoeven et al. Cambridge: Cambridge University Press, pp. 241–264.

Ramanujan, Keerthi and Brenda Weeks (2019). "What is an Akshara?" In: *Handbook of Literacy in Akshara Orthography*. Ed. by R. Malatesha Joshi and Catherine McBride. Switzerland: Springer Nature, pp. 43–54.

Uma Maheshwar Rao, G. (2006). "Analysis of the Telugu corpus: Some preliminary findings". In: *Corpus Linguistics: Study Material for Diploma in Computer Applications in Indian Languages*. Ed. by G. Uma Maheshwara Rao. Hyderabad: University of Hyderabad.

Watson, A. B. (2012). "Perimetric complexity of binary digital images: Notes on calculation and relation to visual complexity". In: *Mathematica Journal* 14, pp. 1–41. DOI: 10.3888/tmj.14-5.

Xiong, J., Y. Zhang, and P. Ju (2021). "The effect of orthographic neighborhood size and the influence of individual differences in linguistic skills during the recognition of Chinese words". In: *Frontiers in Psychology*. DOI: 10.3389/fpsyg.2021.727894.

Yates, Mark, L. Locker, and G. B. Simpson (2004). "The influence of phonological neighborhood on visual word perception". In: *Psychonomic Bulletin and Review* 11, pp. 452–457.

Zhao, J., Q-L. Li, and H-Y. Bi (2012). "The characteristics of Chinese orthographic neighborhood size effect for developing readers". In: *PLoS ONE*. DOI: 10.1371/journal.pone.0046922.