E-ISSN: 1116 - 2619 P-ISSN: 3026-8524
MPRI-JHLAR | INTERNATIONAL JOURNAL OF HUMANITIES, LITERATURE AND ART RESEARCH
VOL. 14 NO. 6 – OCTOBER, 2026

DISCOURSE MARKERS IN HAUSA WRITING: RULES OF WORD COMBINATION AND SEPARATION IN MODERN HAUSA LANGUAGE

SULEIMAN UMAR; & OGAH MUHAMMED IBRAHIM

School Of General And Remedial Studies, Isa Mustapha Agwai Polytechnic Lafia, Nasarawa State, Nigeria.

Corresponding Author: suleimanumar78@gmail.com

DOI: https://doi.org/10.70382/mejhlar.v14i6.0144

Abstract

This study examines discourse markers in contemporary Hausa writing, focusing on the orthographic rules governing their combination (hadewa) and separation (rabewa). It aims to document the structural types of markers in use, quantify combined and separated forms, compare usage with codified conventions, and derive evidence-based standardization guidelines. The study is descriptive and qualitative, supported by quantitative frequency analysis, and draws on Schiffrin's (1987) and Fraser's (1999) frameworks for identifying discourse markers and Haugen's (1966) model of language standardization. It plans a purposive sample of twenty to thirty texts across educational, print/broadcast media, literary, and digital domains. This article reports a preliminary corpus of nine texts (two novels, six newspaper articles, and one linguistics textbook) containing 33 coded occurrences. Of 24 occurrences with a determinate orthographic status, 19 (79.2%) are written separated and five (20.8%) combined. Only five occurrences could be checked against a directly verified rule (Sa'id, 2019): three conform and two deviate (both ko da yaushe written fused, in one novel). The remainder await page-level checks against other authorities. The four Educational occurrences are author-constructed examples from a single textbook. The findings are preliminary and do not characterize Hausa writing generally. Separated forms predominate in this small, unbalanced sample. Completing the corpus, especially digital material and educational texts with running prose, and checking the remaining authorities are necessary before standardization guidelines can be proposed.

Keywords: Discourse markers; Hadewa (Combination); Rabewa (Separation); Hausa orthography; Language standardization

Introduction

Hausa is one of the most widely spoken languages in West Africa, functioning as a major language of intra- and inter-ethnic communication in Nigeria and neighbouring countries (Newman, 2000). Its written tradition has evolved considerably: from Ajami, an Arabic-based script associated historically with Islamic scholarly writing, to Boko, the Roman-based

EDITOR-IN-CHIEF
Prof. S. T. Addo [GHANA]

orthography that now serves as the standard medium for education, print and broadcast media, and official communication (Yahaya, 1988). This transition has been accompanied by sustained efforts to standardize Hausa spelling, word boundaries, and grammatical conventions efforts that remain ongoing rather than fully settled.

Orthographic standardization is of particular importance for a language of Hausa's scale of use: consistent spelling and word-boundary conventions support literacy instruction, editorial practice, and the mutual intelligibility of written texts across regions and media (Amfani et al., 2011). Despite decades of orthographic guidance, however, certain areas of Hausa writing appear to remain unsettled in practice, among them the treatment of discourse markers.

Discourse markers are words and expressions such as conjunctions, connectors, and particles that signal relationships between clauses or sentences (for example, addition, contrast, causation, or sequence) and contribute to the coherence and cohesion of a text without altering its core propositional content (Fraser, 1999; Schiffrin, 1987). Hausa possesses a range of such markers, including multi-word expressions such as saboda haka ("therefore"), ko da yake ("although"), and don haka ("so/therefore"), as well as enclitic particles such as kuwa, fa, and dai (illustrative examples; not empirical corpus data). These markers guide the reader's interpretation of textual relationships and are consequently significant to the coherence of written Hausa.

A specific orthographic question attaches to markers of this kind: whether they should be written as a single combined unit (hadewa) or as separate words (rabewa). Existing orthographic authorities (Amfani et al., 2011; Bunza, 2002; Sa'id, 2019; Yahaya, 1988) provide general rules governing word combination and separation in Hausa, including guidance relevant to particles, prepositions, and conjunctions. Even so, discourse markers as a distinct functional category do not appear to receive systematic, focused treatment in this literature, and current practice as observed informally across educational materials, published texts, and everyday writing suggests that combination and separation may not always be applied consistently. For instance, a form such as saboda haka may in principle be written as two words in keeping with established convention, or rendered as a single fused word, sabodahaka (illustrative example; not empirical corpus data). Which of these reflects, or should reflect, standard practice and how consistently either form is actually used is a question this study is designed to examine empirically rather than assume in advance.

Contemporary digital writing provides an important additional context in which such variation may be observed and investigated. Hausa is used extensively on platforms such as Facebook, WhatsApp, X (formerly Twitter), and YouTube, as well as in online news publishing, and prior scholarship on Hausa digital communication has noted a general tendency toward orthographic variation in these settings (Abbas, 2019; Shuaibu, 2025). Whether, and to what extent, this variation specifically involves the combination and separation of discourse markers is, however, a matter for the empirical investigation undertaken in this study rather than a conclusion that can be drawn in advance of data collection.

This lack of settled guidance poses a practical difficulty for students, teachers, editors, and writers, who must decide which orthographic form to treat as correct in the absence of a category-specific standard. This article responds to that difficulty. It sets out, in the sections that follow, the problem the study addresses (Statement of the Problem), its objectives (Objectives of the Study) and

research questions (Research Questions), a review of relevant conceptual, theoretical, and empirical literature (Review of Related Literature), the theoretical framework guiding the analysis (Theoretical Framework), and the methodology through which the study's discourse-marker corpus is being compiled and analysed (Methodology). The preliminary Stage 2 data and findings are then presented and discussed in the sections from Data Presentation and Analysis through Limitations of the Study.

Statement of the Problem

Contemporary Hausa writing appears to exhibit variation in how discourse markers are represented orthographically, a pattern that, if confirmed and documented, would have implications for students, teachers, writers, and editors who rely on a single accepted form when producing or evaluating written Hausa text. Existing orthographic guidance (Amfani et al., 2011; Bunza, 2002; Sa'id, 2019; Yahaya, 1988) offers general principles for word combination and separation in Hausa but does not appear to address discourse markers as a specifically defined category, nor does it appear to have been tested systematically against usage across the contemporary text types shaped by modernization and digital communication.

This gap matters for several audiences. For students and teachers, an unclear standard complicates instruction in Hausa orthography and the assessment of written work. For writers and editors, it creates uncertainty about which form to adopt or correct. For researchers and language-planning bodies, it represents an area in which codification (Haugen, 1966) the fixing of a norm in reference materials and educational practice does not appear to have been extended to this category of word.

Addressing this problem requires more than restating existing orthographic rules; it requires a systematic, evidence-based comparison of actual contemporary usage against those established rules. The study therefore adopts a descriptive and empirical approach: it is designed to collect naturally occurring examples of discourse markers from a range of contemporary written sources, classify their orthographic forms (combined or separated), and compare these patterns against existing Hausa orthographic conventions in order to identify where usage aligns with, or departs from, established norms. This comparative process is intended to form the evidentiary basis for the study's eventual recommendations. The specific extent and distribution of any inconsistency is a matter for empirical determination through the data collection and analysis described in the Methodology section and is not claimed in advance of that analysis.

Objectives of the Study

This study is guided by the following objectives:

1. To compile a documented inventory of discourse markers occurring in a purposively selected sample of contemporary Hausa written texts, drawn from textbooks, newspapers, literary works, and online/digital platforms.
2. To classify the identified discourse markers according to their structural type (single-word, multi-word, or enclitic) and to quantify the frequency of combined versus separated orthographic forms found in the sampled texts.
3. To compare the documented usage patterns against established Hausa orthographic rules (Amfani et al., 2011; Bunza, 2002; Sa'id, 2019; Yahaya, 1988), identifying specific points of conformity and deviation.

EDITOR-IN-CHIEF
Prof. S. T. Addo [GHANA]

4. To propose a set of evidence-based, standardized orthographic guidelines for the combination and separation of discourse markers, derived from the study's findings, for use by educators, editors, and language-planning bodies.

Research Questions

Each research question below corresponds directly to one of the four objectives set out in the Objectives of the Study section:

RQ1: What discourse markers occur in contemporary Hausa written texts across the sampled educational, print/broadcast media, literary, and digital domains, and how are they structurally classifiable (single-word, multi-word, or enclitic)?

RQ2: What is the frequency distribution of combined versus separated orthographic forms for each type of discourse marker identified?

RQ3: To what extent does documented usage conform to or deviate from established Hausa orthographic conventions, and in which text types is deviation most concentrated?

RQ4: What standardized combination and separation guidelines can be derived from the documented usage patterns for educators, editors, and language-planning bodies?

Conceptual Clarification

A discourse marker refers to a word or expression such as a conjunction, connector, or particle that signals relationships between clauses or sentences (e.g., addition, contrast, causation, sequence) and contributes to the coherence and cohesion of a text without altering its core propositional meaning (Fraser, 1999; Schiffrin, 1987; cf. Crystal, 2008, for a general linguistic definition of discourse and connective particles). In the Hausa grammatical tradition, the concepts this study investigates are more specifically referred to as hadewa (combination) and rabewa (separation), terms used by Hausa orthographers to describe whether a word or word-group is written as one continuous unit or as two or more separate units (Bunza, 2002; Sa'id, 2019; Yahaya, 1988). It should be noted that hadewa and rabewa are descriptive categories rather than a presumption that any given multi-word marker ought to be combined, or any given particle ought to be separated; the appropriate orthographic treatment of each marker is precisely what this study investigates, rather than an assumption built into its design.

Modern Hausa writing, for the purposes of this study, refers to written Hausa produced using the standard Boko (Roman-script) orthography across contemporary domains education, print and broadcast media, literature, and digital communication reflecting current usage rather than historical or Ajami-script conventions.

Discourse Markers

Schiffrin (1987) conceptualizes discourse markers as linguistic elements operating at the boundaries of talk or text, functioning to signal relationships between adjacent discourse units and contributing to overall textual coherence. Fraser (1999) refines this view, defining discourse markers as a pragmatically distinct class of expressions including conjunctions, adverbials, and particles that signal a specific relationship between the segment they introduce and the segment preceding it, without altering the truth-conditional meaning of either. These two accounts, taken

together, provide the general analytical vocabulary structural type and discourse function applied to Hausa data later in this study.

Hausa Orthography and Word Boundaries

Yahaya (1988) contributes foundational rules of word separation and combination in Hausa, particularly with regard to particles, prepositions, and conjunctions. Bunza (2002) similarly addresses the writing conventions governing standard Hausa orthography. Sa'id (2019) provides a more recent, pedagogically oriented treatment of the same rules, explicitly framed as a guide for writers (Jagora ga Marubuta). Of particular relevance to the present study, Sa'id (2019, p. 10) states that connective/general words (kalmomin sadarwa) are ordinarily written as a single combined word giving koyaushe, kowane, watakila, wannan, and yakamata, among others, as examples but specifies that certain of these forms must instead be written separated once a change occurs in their structure, giving the combined koyaushe versus the separated ko da yaushe as an explicit example, alongside abin da, lokacin da, in ban da, tun da, and har da. This is among the few instances in the reviewed literature where a structural condition for separating an otherwise-combined connective word is stated explicitly, rather than left to be inferred from general principle. Amfani et al. (2011) present a pan-regional reference on standardized Hausa orthography, developed collaboratively across Nigeria, Niger, Cameroon, Ghana, and Chad. Together, these four works constitute the principal set of orthographic authorities against which this study's documented usage patterns will be compared. None of the four, however, is understood to treat discourse markers as a comprehensively and separately defined orthographic category, which is the specific gap this study addresses (see the Research Gap subsection).

Discourse Markers in Hausa

Newman (2000) provides a comprehensive grammatical description of Hausa, discussing function words and particles that often operate as discourse markers, although without a focused treatment of their orthographic behaviour. Jaggar (2001) examines Hausa syntax and semantics, highlighting the functional roles of connectors and particles in structuring discourse. More directly relevant to the present study's category of interest, Kiyawa (2021) analyses discourse markers within a corpus of thirty-six Hausa proverbs, identifying their interpersonal and cohesive functions in traditional, orally derived texts. Zając (2024) offers a detailed pragmatic analysis of the Hausa marker to, demonstrating that individual Hausa discourse/pragmatic markers continue to attract focused linguistic attention, though consistent with the gap this study addresses from a functional-semantic rather than orthographic perspective.

Empirical Studies on Hausa Orthographic Variation

A smaller body of empirical work has begun to document patterns of Hausa orthographic usage in digital contexts. Abbas (2019) conducted a synchronic study of Hausa Facebook users' writing, reporting that some users represent their names using graphemes not native to standard Hausa orthography; this finding concerns non-standard grapheme substitution rather than discourse-marker combination or separation specifically, and is cited here only as general evidence of non-standard/deviant orthographic representation among Hausa Facebook users. Shuaibu (2025)

analyses grammatical and orthographic errors in Hausa Facebook writing, describing a pattern of what the study terms "destandardisation" in digital Hausa vernacular, in which conventional word-boundary rules are frequently not observed; this is cited for its account of Hausa Facebook writing and digital destandardisation generally, and not as a direct study of discourse-marker combination or separation. Abdullahi et al. (2024) apply computational sentiment-analysis methods to abbreviated and acronym forms in Hausa Twitter data, demonstrating that irregular Hausa digital-text forms have attracted attention from natural-language-processing research; this is cited only for its treatment of Hausa abbreviations and computational digital-text analysis, and not as direct evidence about discourse-marker orthography.

Taken together, these studies indicate that Hausa orthographic variation, including in digital communication, has attracted growing empirical attention, and that individual Hausa discourse markers have received some focused linguistic attention (e.g., Kiyawa, 2021; Zając, 2024). However, the literature reviewed for this study does not appear to include a study that specifically and systematically examines the combination and separation patterns of discourse markers as a defined category, across multiple text types (educational, print/broadcast media, literary, and digital), and compares these patterns against established orthographic rules.

Research Gap

The literature reviewed does not appear to contain a systematic study that specifically examines the combination and separation of discourse markers as a defined orthographic category across educational, print/broadcast, literary, and digital Hausa texts while comparing documented usage against established orthographic conventions (Amfani et al., 2011; Bunza, 2002; Sa'id, 2019; Yahaya, 1988). This is presented as a gap in the literature reviewed for this study, not as a claim that no relevant work exists anywhere in Hausa scholarship. The present study responds to this gap by proposing a focused, data-driven, empirical account of discourse-marker combination and separation across a range of contemporary Hausa written sources.

Theoretical Framework

This study is anchored on two complementary theoretical frameworks: Schiffrin's (1987) discourse-coherence approach to discourse markers, as extended by Fraser's (1999) grammatical-pragmatic account, and Haugen's (1966) language planning model.

Schiffrin (1987) and Fraser (1999) together provide the analytical foundation for identifying and classifying discourse markers by structural type (single-word, multi-word, enclitic) and by discourse function (contrast, cause, addition, sequencing). This pairing directly supports Objectives One and Two, which require an inventory and structural classification of the discourse markers found in the sampled texts.

Haugen's (1966) language planning model provides the study's standardization lens. Haugen identifies four major processes in language standardization: (1) selection of a linguistic norm, (2) codification of that norm including orthographic and grammatical standardization, (3) implementation of the codified norm through education and media, and (4) elaboration of the norm's functional range to meet evolving communicative needs. This study is situated primarily within the codification stage: by documenting current usage patterns of discourse-marker

EDITOR-IN-CHIEF
Prof. S. T. Addo [GHANA]

combination and separation and comparing them against existing orthographic conventions (Amfani et al., 2011; Bunza, 2002; Sa'id, 2019; Yahaya, 1988), the study is intended to contribute to this stage and to lay groundwork for the subsequent implementation stage, typically led by educational and language-planning bodies. This is offered as a modest, specific contribution to one stage of Haugen's model rather than a claim to resolve Hausa orthographic standardization as a whole.

Together, these two frameworks allow the study to address both the pragmatic-functional dimension of discourse markers (how and why they operate in text) and the orthographic-standardization dimension (how their combination and separation might be codified), supporting all four of the study's objectives.

Methodology

The description below sets out the research design and procedures for the full study, distinguishing work already completed from work still planned. A preliminary Stage 2 corpus has been compiled: nine clean texts (the two novels Ya'yan Safara and Bakin Suri, six Aminiya articles, and one textbook, Mukhtar's (2024) An Introductory Hausa Linguistics) have been read and coded, producing 33 coded discourse-marker occurrences. Work still planned, and therefore described in the future tense, comprises expanding the corpus toward the planned 20–30 texts; acquiring and analysing additional eligible Educational and Digital/Online material; checking the page-level provisions of Yahaya (1988), Bunza (2002), and Amfani et al. (2011) against the coded occurrences; and conducting the focus-group validation exercise.

Research Design

The study is designed around a descriptive and empirical qualitative research approach (Creswell, 2014), suited to examining naturally occurring language data and identifying patterns of orthographic usage in written texts. Quantitative elements, such as frequency counts, have been applied to the preliminary Stage 2 corpus and will be extended to the fuller corpus as it is compiled.

Study Area and Data Sources

Data will be drawn from written Hausa texts relevant to the Nigerian research context, covering the period 2000–2025. For the educational, print/broadcast media, and literary categories, this means material produced and published within Nigeria, particularly in the major Hausa-speaking states of Kano, Katsina, Sokoto, Kaduna, and Nasarawa. Four categories of sources are envisaged: (1) educational texts secondary and tertiary-level Hausa language textbooks approved for use in Nigerian schools, produced and published within Nigeria; (2) print and broadcast media Hausa-language newspapers, principally Aminiya (Daily Trust group), and broadcast archival materials, produced and published within Nigeria; (3) literary works contemporary Hausa prose fiction (Littattafan Soyayya), produced and published within Nigeria; and (4) digital/online media publicly available Hausa-language material meeting the criteria set out below. Because digital/online material may be hosted or maintained outside Nigeria even where it is written in Hausa and otherwise relevant to this study, sources in this fourth category are not required to have been produced or published within Nigeria in the sense that applies to the first three; no digital source

EDITOR-IN-CHIEF
Prof. S. T. Addo [GHANA]

is described as Nigerian in origin unless this can actually be established, and each is instead assessed individually against the criteria below.

The literary corpus actually compiled and coded consists of two novels genuinely acquired as physical copies and read: Ya'yan Safara by Lamido (2014) and Bakin Suri by Amin (2011), both works of contemporary Hausa prose fiction within the Littattafan Soyayya tradition. Any further literary source added at a later stage must remain a genuine, published work of contemporary Hausa prose fiction, acquired in a form that allows its text to be read and coded, subject to verification of the acquired copy against its bibliographic details.

The educational text actually compiled and coded is Mukhtar's (2024) An Introductory Hausa Linguistics (Bayero University Press, Kano), which is a tertiary-level, Nigeria-published, in-period Hausa linguistics textbook.

The digital/online source was originally scoped to Facebook, WhatsApp public group archives, X (formerly Twitter), YouTube comments, and online Hausa news platforms. Wikipedia was considered as a prospective additional source at an intermediate stage but was not ultimately included: on audit, no Wikipedia candidate could be shown to satisfy criterion (e) below, and none was retained in the corpus. The only digital source actually acquired and read to date is a Hausa-language page from Hesperian Health Guides (n.d.; ha.hesperian.org), itself a Hausa translation of the English-language health manual "Where There Is No Doctor."

Population and Sampling

The population comprises written Hausa texts authored or produced by Hausa-speaking writers, journalists, textbook authors, and content creators. For the educational, print/broadcast media, and literary categories, the population is bounded to material published or circulated within Nigeria between 2000 and 2025; for the digital/online category, the population instead comprises material satisfying the criteria set out in the Study Area and Data Sources subsection, since such material may be hosted outside Nigeria even where it is otherwise relevant to this study. This population is deliberately bounded to Nigeria-published material for the first three categories, and to material meeting the criteria in the Study Area and Data Sources subsection for the fourth rather than extended to the wider global Hausa-speaking diaspora without qualification, in order to ensure feasibility of data collection. A purposive sampling technique has been used to select the texts compiled so far and will continue to be used to select further texts that are relevant to the study and likely to contain identifiable occurrences of discourse markers. The study targets a sample of 20–30 texts, distributed proportionately across the four categories (approximately 5–8 texts per category); this target distribution is unaffected by the source revisions described in the Study Area and Data Sources subsection, since both remain single categories within the same four-category design. The preliminary Stage 2 corpus comprises nine clean texts (two Literary, six Print/Broadcast Media, and one Educational) and no Digital/Online texts; it is therefore below the planned target and unevenly distributed across categories.

Data Collection Procedure

Print and archival materials were collected for the preliminary Stage 2 corpus as follows: the two novels, Ya'yan Safara and Bakin Suri, were acquired as physical copies and read, the six Aminiya

EDITOR-IN-CHIEF
Prof. S. T. Addo [GHANA]

articles (Aminiya, 2000–2025) were sourced from Arewa house bookshop in Kaduna, and the textbook An Introductory Hausa Linguistics was acquired as a physical copy from Bayero University Kano, of which pages 90–92 were acquired and read. Digital texts will be collected through manual review and text-capture of publicly available material from the platforms named in the Study Area and Data Sources subsection Facebook, WhatsApp public group archives, X, YouTube comments, and online Hausa news platforms where a specific item satisfies criteria (a)–(g) set out there (written in contemporary Hausa; publicly accessible; sufficient continuous prose; identifiable discourse-marker occurrences; a verified fit within the 2000–2025 period; adequate source documentation; and, for translated/adapted material, confirmed eligibility under an explicit translation-status decision), with attention to preserving the surrounding context needed for discourse analysis. In the preliminary stage, the texts were read systematically, and instances of discourse markers were identified and extracted; each instance was documented in a data sheet (the Stage 2 extraction workbook) indicating the source (with full title, author, and publication details), context, and orthographic form (combined or separated). The same procedure will be applied to further texts, and any digital source added under the criteria in the Study Area and Data Sources subsection will be documented on the same basis, so that its eligibility can be checked independently rather than assumed.

The original proposal also envisages scholarly consultation, including focus-group discussions with Hausa language scholars, to validate the classification of collected data. The preliminary findings and classification of the collected data were reviewed through scholarly consultations with Hausa language professors. These consultations provided expert feedback on the interpretation and classification of the discourse markers. The findings were revised where necessary in light of the feedback received.

Because part of the corpus will be drawn from publicly available digital communication, data collection will need to attend to the ethical handling of such material for example, by anonymizing usernames or other identifying information in any digital example quoted or reproduced in subsequent reporting. Ethical handling of such material, including any required institutional approval, will be documented when digital data collection begins.

Data Coding and Analysis

Data analysis involves qualitative content analysis (Krippendorff, 2018) supported by simple quantitative measures. Each documented discourse-marker occurrence in the preliminary corpus was coded for the following fields, which will also be applied to any further texts: (1) identification number; (2) discourse marker; (3) structural type (single-word, multi-word, or enclitic); (4) discourse function; (5) observed orthographic form; (6) combined/separated status; (7) source category; (8) source title; (9) author or content creator, where applicable; (10) publication or post date; (11) page or post location, where available; (12) surrounding context; (13) the relevant established orthographic rule against which the form is assessed; (14) an assessment of conformity or deviation; and (15) analytical notes. A marker that is merely mentioned (for example, named in a grammatical description or listed as an example word) was not coded as an occurrence; a marker used within an author-constructed example sentence was coded, with that status recorded in the analytical notes.

Frequency counts of combined versus separated forms were manually coded and tallied for the preliminary corpus using a structured spreadsheet to identify dominant patterns of combination and separation. Where a directly checked rule was available, the observed forms were compared with it to record conformity or deviation; at present this was possible only for Sa'id (2019). Comparison with the page-level provisions of Yahaya (1988), Bunza (2002), and Amfani et al. (2011) has not yet been carried out and remains planned. Preliminary numerical results, based on the below-target corpus compiled to date, the coding framework above was applied to produce them and will continue to guide coding as the remaining planned corpus is compiled.

Data Presentation and Analysis

This section presents the discourse-marker occurrences documented in the Stage 2 corpus to date. All figures reported here are drawn only from occurrences coded as verified in the Stage 2 extraction workbook and matched to a documented source record; occurrences drawn from Sa'id's (2019) orthography manual and mentions of markers in Mukhtar's (2024) textbook that are not uses of them (a prose mention and a bare word list), occurrences whose underlying source could not be matched to a documented source record, and one marker whose identity remains unconfirmed ('hakake') are excluded from the counts and are discussed separately. On this basis, the clean dataset comprises 33 discourse-marker occurrences: 19 from the Literary category (Ya'yan Safara [Lamido, 2014] and Bakin Suri [Amin, 2011]), 10 from the Print/Broadcast Media category (six Aminiya articles), and 4 from the Educational category (one tertiary-level textbook, An Introductory Hausa Linguistics by Mukhtar). No occurrence from the Digital/Online category currently meets the criteria for inclusion in this clean set.

Inventory of Discourse Markers

Table 1 lists every distinct discourse marker documented in the clean dataset, with its structural type, discourse function, the source category or categories in which it was found, and its total number of occurrences. Twenty-one distinct markers are represented, ranging from single occurrences to a maximum of four (saboda haka and amma).

Table 1: Inventory of Discourse Markers (Clean Corpus, n = 33 occurrences)

No.	Marker	Structural Type	Discourse Function	Source Category	Occurrences
1	saboda haka	Multi-word	Cause/Result ("therefore")	Literary; Print/Broadcast Media	4
2	amma	Single-word	Contrastive ("but")	Literary; Print/Broadcast Media; Educational	4
3	amma fa	Multi-word (two single-word markers in sequence)	Contrast + emphasis	Literary	2
4	sai dai	Multi-word	Contrast/qualification	Print/Broadcast Media	2
5	don haka	Multi-word	Cause/Result ("so/therefore")	Literary; Print/Broadcast Media	2
6	ko da yaushe	Multi-word	Temporal/frequency ("always, at any time")	Literary	2

EDITOR-IN-CHIEF
Prof. S. T. Addo [GHANA]

No.	Marker	Structural Type	Discourse Function	Source Category	Occurrences
7	fa	Enclitic	Emphatic/attention-drawing particle	Literary; Print/Broadcast Media	2
8	sannan	Single-word	Sequential ("then/and then")	Literary; Educational	2
9	ai	Single-word	Emphatic ("indeed/well")	Literary	1
10	dai	Enclitic	Emphatic/contrastive-softening particle	Literary	1
11	daga nan	Multi-word	Sequential ("from there/then")	Literary	1
12	bugu da kari	Multi-word	Additive ("furthermore/moreover")	Literary	1
13	hakazalika	Single-word	Additive/parallel	Print/Broadcast Media	1
14	duk da haka	Multi-word	Concessive/contrastive	Print/Broadcast Media	1
15	don kuwa	Multi-word	Cause/Result ("because/for")	Literary	1
16	ko da yake	Multi-word	Concessive	Print/Broadcast Media	1
17	saboda	Single-word	Cause ("because of/due to")	Literary	1
18	kowace	Single-word	Universal/distributive ("each/every")	Literary	1
19	sai dai kuma	Multi-word	Contrast/qualification	Print/Broadcast Media	1
20	kuma	Single-word	Additive ("and also")	Educational	1
21	watakila	Single-word	Epistemic/hedging ("perhaps")	Educational	1

Structural Classification

Of the 33 occurrences, 18 are multi-word (e.g., saboda haka, don haka, ko da yaushe, sai dai kuma, duk da haka, don kuwa, ko da yake, daga nan, bugu da kari, and the two amma fa cases, each coded as a multi-word sequence of two single-word markers), 3 are enclitic (fa ×2, dai ×1), and 12 are single-word (amma ×4, hakazalika, ai, sannan ×2, kuma, saboda, kowace, watakila); of these 12, 9 (amma ×4, hakazalika, ai, sannan ×2, kuma) carry no combination/separation status, while the remaining 3 (saboda, kowace, watakila) do carry a determinate status (all Combined, per Table 2), since Sa'id (2019, p. 10, Rule R1) treats them as connective words that are correctly written as one word in their base form. This distribution reflects the corpus as coded rather than a designed quota, and should not be read as evidence of the true proportions of marker types in Hausa writing generally, given the corpus's small size and its concentration in two literary texts and six newspaper articles, with a single textbook contributing four constructed examples.

Combined and Separated Orthographic Forms

Table 2 reports the 24 occurrences in the clean dataset that carry a determinate combined/separated status, across 16 distinct markers. These 16 include all multi-word and enclitic markers, together with the three single-word forms saboda, kowace, and watakila, since Sa'id (2019, p. 10, Rule R1) treats these three as connective words with a fixed, rule-governed combined form; a determinate status therefore applies to them just as it does to the multi-word and enclitic markers. The remaining nine single-word occurrences (amma ×4, hakazalika, ai, sannan ×2, kuma) are excluded from Table 2, and are marked 'n/a' in Table 4 (Table 3 lists verified judgements only),

because no combination/separation question applies to them at all; they are retained in Table 1's inventory but do not enter this table's counts. Of the 24 determinate occurrences, 5 are written combined (20.8%) and 19 are written separated (79.2%).

Table 2: Combined and Separated Forms (Clean Corpus)

Marker	Combined	Separated	Total	% Combined	% Separated
saboda haka	0	4	4	0.0%	100.0%
amma fa	0	2	2	0.0%	100.0%
sai dai	0	2	2	0.0%	100.0%
don haka	0	2	2	0.0%	100.0%
ko da yaushe	2	0	2	100.0%	0.0%
fa	0	2	2	0.0%	100.0%
daga nan	0	1	1	0.0%	100.0%
bugu da kari	0	1	1	0.0%	100.0%
duk da haka	0	1	1	0.0%	100.0%
don kuwa	0	1	1	0.0%	100.0%
ko da yake	0	1	1	0.0%	100.0%
saboda	1	0	1	100.0%	0.0%
kowace	1	0	1	100.0%	0.0%
watakila	1	0	1	100.0%	0.0%
sai dai kuma	0	1	1	0.0%	100.0%
dai	0	1	1	0.0%	100.0%
TOTAL	5	19	24	20.8%	79.2%

The three 'Combined' markers with a determinate status other than ko da yaushe (saboda, kowace, watakila) are simple single-word forms that Sa'id (2019, p. 10) records as inherently written combined; their 'Combined' status reflects observed usage that happens to match that baseline form, not a separation question resolved in favour of combination. In the case of watakila, the observed form comes from an author-constructed example sentence (Distribution Across Text Categories subsection), so its 'Combined' status records the textbook author's own spelling in an illustrative sentence rather than a writer's choice in running text.

Conformity and Deviation from Established Rules

A conformity or deviation judgement is recorded only where an identifiable, directly-checked rule from an orthographic authority applies to the observed form; all other occurrences are recorded as 'Unclear/Needs review' rather than assigned a judgement the evidence does not support. On this basis, only five of the 33 occurrences currently carry a Conforms/Deviates judgement, all resting on Sa'id (2019, p. 10, Rule R1), the only orthographic-authority rule directly and photographically verified against the primary text during Stage 2: two occurrences of the fused form 'kodayaushe'/'Kodayaushe' (Ya'yan Safara, pp. 17 and 20) are judged to Deviate, since Rule R1 explicitly requires 'ko da yaushe' to be written separated once 'da' is present; and three occurrences ('kowace' and 'saboda', Bakin Suri, pp. 22 and 20; and 'watakila', from Mukhtar's (2024) textbook,) are judged to Conform, matching Rule R1's list of connective words that are correctly written combined. The remaining 28 occurrences carry no conformity/deviation

judgement: 19 are recorded as 'Unclear/Needs review', while 9 are recorded as 'n/a' because they are single-word markers (amma ×4, hakazalika, ai, sannan ×2, kuma) for which a combination/separation judgement does not apply. Among the 19 Unclear/Needs-review cases, the general treatments in Yahaya (1988), Bunza (2002), and Amfani et al. (2011) are understood to bear on markers such as saboda haka, don haka, sai dai, and the enclitics fa and dai, but the specific page-level provisions have not yet been directly checked against the primary texts in the way Sa'id's (2019) p. 10 has been, so no conformity judgement is asserted for them at this stage. Table 3 lists only the five verified judgements; the 19 Unclear/Needs-review and nine n/a occurrences do not appear in it.

Table 3: Conformity with Established Orthographic Rules (verified judgements only)

Marker	Observed Form	Established Form	Authority	Judgement	Category	ID
ko da yaushe	kodayaushe (fused)	ko da yaushe (separated)	Sa'id, 2019, p. 10 (R1)	Deviates	Literary	L1-1
ko da yaushe	Kodayaushe (fused)	ko da yaushe (separated)	Sa'id, 2019, p. 10 (R1)	Deviates	Literary	L1-2
kowace	kowace (combined)	kowace (combined)	Sa'id, 2019, p. 10 (R1)	Conforms	Literary	L2-12
saboda	saboda (combined)	saboda (combined)	Sa'id, 2019, p. 10 (R1)	Conforms	Literary	L2-14
watakila	watakila (combined)	watakila (combined)	Sa'id, 2019, p. 10 (R1)	Conforms	Educational	E3-4

Distribution Across Text Categories

Table 4 distributes the 33 occurrences across the four source categories. It reports two different kinds of classification side by side, and they should not be read as measuring the same thing. The Combined and Separated columns record the observed orthographic form of the 24 occurrences with a determinate status; the nine n/a occurrences have no such status. The Conforms, Deviates, Unclear, and n/a columns record whether the observed form could be judged against a directly checked rule, and together account for all 33 occurrences. A separated form is therefore not by itself a conforming or deviating form: all 19 separated occurrences are currently Unclear/Needs review, whereas the five Combined occurrences are those judged against Sa'id (2019, p. 10, Rule R1).

Table 4: Distribution by Corpus Category

Category	Texts	Occurrences	Combined	Separated	Conforms	Deviates	Unclear	n/a
Educational	1	4	1	0	1	0	0	3
Print/Broadcast Media	6	10	0	8	0	0	8	2
Literary	2	19	4	11	2	2	11	4
Digital/Online	0 (usable)	0	0	0	0	0	0	0
TOTAL	9	33	5	19	3	2	19	9

The Digital/Online row is zero because no eligible occurrence from that category is included, not because the phenomenon is absent from it. The Educational row rests on a single textbook and on

four occurrences drawn from one author-constructed example set (p. 92, example 83) rather than running prose, so it cannot be read as evidence of how Educational texts in general write these markers. The three populated categories differ greatly in the number of texts and occurrences and should not be compared as if they were balanced samples.

Discussion of Findings

The clean Stage 2 dataset contains mostly separated forms: 19 of 24 determinate occurrences (79.2%) are written as separate words, against 5 combined (20.8%). This describes the observed forms only and is not a finding about conformity: the 19 separated occurrences all remain Unclear/Needs review because the page-level provisions of Yahaya (1988), Bunza (2002), and Amfani et al. (2011) have not yet been checked against them. Nor does the pattern establish a general preference in Hausa writing, given the small and unbalanced corpus.

The only directly-verified rule applied so far, Sa'id's (2019, p. 10) Rule R1, produces a more mixed picture: three occurrences conform ('kowace', 'saboda', and 'watakila', the last from a constructed example in a textbook) and two deviate (both instances of 'kodayaushe'/'Kodayaushe' for 'ko da yaushe'), both from the same novel, Ya'yan Safara. Sa'id's rule is notable among the literature reviewed for this study in stating explicitly that a connective word which is ordinarily combined (koyaushe) must be written separated once its form changes through the insertion of 'da' (ko da yaushe); how Bunza (2002) and Amfani et al. (2011) treat this distinction has not yet been checked at page level. The two deviating instances both come from a single author's text, and a sample of two occurrences from one novel cannot establish whether this reflects an idiosyncrasy of this author, a broader tendency in Littattafan Soyayya prose, or neither.

Viewed through Haugen's (1966) codification stage, the two verified deviation cases indicate at least one concrete point at which a codified rule (Sa'id's [2019, p. 10] exception for 'ko da yaushe') is not being followed in at least one contemporary literary text, which is the kind of documented gap this study's codification-stage contribution is intended to surface. The much larger set of Unclear/Needs-review occurrences, however, means that no general claim about Hausa writers' conformity or deviation as a whole can yet be made; the corpus compiled to date supports specific, narrow observations rather than the broad characterizations the literature review flagged as absent from prior scholarship (Research Gap subsection).

The corpus's thin Educational material (one textbook contributing four occurrences) and its lack of verified Digital/Online material limit what can be said about either category. The Educational occurrences in particular should not be read as evidence of how educational writers use these markers: they are alternatives within one illustrative sentence composed by the textbook's author to demonstrate coordination, and they show only that the author spelled the one rule-governed form among them ('watakila') in the combined form that Rule R1 prescribes. The single Digital/Online source located and read so far (a Hausa translation of an English-language health guide) is, in addition, translated rather than originally-authored Hausa material, a distinction this study's methodology does not yet explicitly resolve (Study Area and Data Sources subsection); the empirical account of digital Hausa orthographic variation offered by Abbas (2019) and Shuaibu (2025) therefore cannot yet be compared against this study's own digital data, and this comparison remains for a later stage of data collection.

EDITOR-IN-CHIEF
Prof. S. T. Addo [GHANA]

Hausa Orthography

The two verified deviation cases for 'ko da yaushe' indicate that at least one codified exception in Sa'id's (2019) guide is not being consistently followed in contemporary literary prose, suggesting this specific form may merit attention in any future consolidation or cross-referencing of Hausa orthographic authorities.

Teaching and Learning

Where a specific, documented rule such as Sa'id's (2019) treatment of 'ko da yaushe' exists, it offers teachers a concrete, citable example for illustrating how word-boundary rules can depend on a marker's internal structure; the study does not yet have enough verified rule-checked cases to support broader pedagogical claims beyond this one example.

Editing and Publishing

For editors of Littattafan Soyayya fiction specifically, the recurrence of the fused spelling across two separate pages of the same novel suggests that a targeted editorial check for this specific form could be worthwhile; this observation is naturally limited to the one novel and one form directly verified here.

Digital Hausa Writing

No implication is drawn for digital Hausa writing at this stage, since the corpus does not yet contain a verified, originally-authored digital source; this is flagged as an area for the next stage of data collection rather than a gap the present findings can fill.

Language Standardization

Within Haugen's (1966) framework, the study's codification-stage contribution so far is narrow but concrete: one documented point of divergence between a codified rule and observed usage. Elaborating this into standardization guidance for language-planning bodies (Objective 4) requires the fuller corpus targeted in the Population and Sampling subsection, and the Recommendations section accordingly frames its recommendations as provisional.

Conclusion

This study set out to document discourse markers in contemporary Hausa writing and examine their orthographic combination and separation against established rules. On the material actually compiled and coded to date 33 occurrences across two literary novels, six Aminiya newspaper articles, and one linguistics textbook separated forms substantially outnumber combined forms (79.2% versus 20.8% of the determinate cases in this small preliminary corpus), and the two occurrences directly checked against a documented, primary-source rule (Sa'id's [2019] treatment of 'ko da yaushe') were both found to deviate from it, while three other occurrences ('kowace', 'saboda', 'watakila') were found to conform to the same source's general combined-form rule. These are the only findings this study can currently support with a directly verified authority rule.The Educational category is represented only by one textbook, and the Digital/Online category is not yet represented by any verified, eligible corpus occurrence. No claim is made here that these

EDITOR-IN-CHIEF
Prof. S. T. Addo [GHANA]

findings characterize Hausa writing generally; they describe, specifically and only, the texts this study has so far obtained and read.

Recommendations

For Writers and Editors

- Treat 'ko da yaushe' as a separated form distinct from the combined 'koyaushe', per Sa'id's (2019, p. 10) documented rule, and check manuscripts for the fused spelling observed twice in this corpus.

For Teachers and Learners

- Use the documented 'koyaushe' / 'ko da yaushe' distinction as a concrete classroom example of how inserting 'da' changes a connective word's required orthographic form.

For Publishers and Media

- Consider a targeted proofreading check for multi-word causal and temporal connectives in locally printed Littattafan Soyayya fiction, where this study's only two verified deviations both occurred.

For Hausa Language-Planning Institutions

- Treat the findings here as provisional evidence supporting further, larger-scale verification of discourse-marker orthography, rather than as a basis for new standardization guidance on their own.

For Future Research

- Expand the corpus toward the planned 20–30 texts, with particular priority on the Digital/Online category, which has no verified, eligible occurrences, and on Educational texts containing running Hausa prose, since the one Educational text so far contributes only constructed examples.
- Directly check the specific page-level provisions of Yahaya (1988), Bunza (2002), and Amfani et al. (2011) against the 19 occurrences currently marked Unclear/Needs review, following the same photograph-and-transcribe procedure used for Sa'id (2019).
- Resolve, as a methodological decision rather than a data question, whether translated/adapted Hausa material (such as the Hesperian health guide located so far) is eligible for the Digital/Online category, and document that decision in the Study Area and Data Sources subsection before further digital sources are added.

These recommendations are the study's own, derived from the findings reported in the Data Presentation and Analysis and Discussion of Findings sections; they are distinguished here from the existing orthographic rules of Yahaya (1988), Bunza (2002), Sa'id (2019), and Amfani et al. (2011), which the recommendations draw on but do not restate as if they were this study's own conclusions.

Limitations of the Study

- Corpus size: ten corpus texts have been acquired and read to date against a planned target of 20–30, but only nine of these the two literary novels, the six Aminiya articles, and the Mukhtar (2024) textbook form the clean dataset behind the sections from Data

Presentation and Analysis through Limitations of the Study (33 occurrences). The tenth text, the Hesperian Hausa-translation page, was read and logged but its single coded occurrence is excluded from the clean dataset pending the methodological decision, described in the Study Area and Data Sources subsection, on whether translated/adapted Hausa material is eligible for the Digital/Online category. The Digital/Online category is not yet represented by any verified, eligible occurrence, the Educational category is represented only by four occurrences from a single textbook, and the findings above should be read as based chiefly on the Literary and Print/Broadcast Media categories. The findings are preliminary and must not be generalized to Hausa writing as a whole.

- Category imbalance: of the 33 clean-dataset occurrences, 19 (57.6%) come from just two literary novels by two authors, 10 (30.3%) from six articles from a single newspaper (Aminiya), and 4 (12.1%) from three pages of a single textbook; the corpus does not yet reflect the breadth of sources the study's design specifies.
- Constructed evidence in the Educational category: the Educational text (Mukhtar, 2024) was admitted by researcher, although the Hausa on the pages read consists of headings, isolated example sentences, and word lists. Its four counted occurrences are alternatives within one author-constructed example sentence, they were verified as connective uses in that sentence, but they are not naturally occurring usage and are not independent observations of writers' choices.
- Source availability: the originally planned literary source, Balaraba Ramat Yakubu, was not available for acquisition; a later proposed substitute, Mace Mutum by Rahma Abdul-Majid, was also not obtained and is not part of the corpus actually compiled. The literary corpus instead consists of Ya'yan Safara (Lamido, 2014) and Bakin Suri (Amin, 2011), acquired as physical copies.
- Digital-source limitations: the only Digital/Online source located and read to date is a Hausa translation of an English-language source (a Hesperian health guide), and the study's methodology does not yet explicitly state whether translated/adapted Hausa material is eligible for this category; this source's single coded occurrence is accordingly excluded from the figures in the Data Presentation and Analysis section pending that methodological decision, and no other candidate digital source (including a Facebook post considered during Stage 2) has yet been logged with a verifiable source record.
- Uncertainty in some forms: one candidate marker, 'hakake' (Bakin Suri, p. 12), was documented but its identity has not been confirmed and is excluded from all counts in the Data Presentation and Analysis section pending clearer verification of the underlying photograph.
- Rule verification: of the four orthographic authorities named in this study's objectives, only Sa'id (2019) has so far been directly checked page-by-page against the corpus; the treatments in Yahaya (1988), Bunza (2002), and Amfani et al. (2011) remain to be checked in the same way, so 28 of the 33 clean occurrences currently carry no conformity judgement.
- Although scholarly consultations were conducted separately with Hausa language professors, a collective focus-group discussion envisaged in the original proposal was not

held. Therefore, the feedback obtained reflects individual expert consultations rather than a collective discussion.

References

Abbas, Y. (2019). Synchronic study of Hausa Facebookers' deviant orthographic representation. *Alsun Journal of Humanities, 2*(1), 99–104.

Abdullahi, H. I., Ahmad, M. A., & Haruna, K. (2024). Twitter sentiment analysis for Hausa abbreviations and acronyms. *Science World Journal, 19*(1), 101–104. https://doi.org/10.4314/swj.v19i1.13

Amfani, A. H., Umar, A., Abdoulaye, M., Attouman, M. B., Malumfashi, A. I., & Munkaila, M. (2011). *A unified standard orthography for the Hausa language (Nigeria, Niger, Cameroon, Ghana and Chad)* (Monograph Series No. 241). Centre for Advanced Studies of African Society.

Amin, M. L. (2011). *Bakin suri*. Ahmadu Bello University Press.

Aminiya. (2000–2025). Hausa weekly newspaper. Media Trust Limited.

Bunza, A. M. (2002). *Rubutun Hausa: Yadda yake da yadda ake yin sa*. Ibrash Islamic Publications Centre.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.

Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Blackwell Publishing.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics, 31*(7), 931–952.

Haugen, E. (1966). Dialect, language, nation. *American Anthropologist, 68*(4), 922–935.

Hesperian Health Guides. (n.d.). *Sabon littafin a inda babu likita* [The new where there is no doctor] (A. Ahmad, Trans.). Hesperian Health Guides Open Publishing. hesperian.org

Jaggar, P. J. (2001). *Hausa*. John Benjamins.

Kiyawa, H. A. (2021). Discourse markers in Hausa proverbs: Exploring intellectual wise sayings from African wisdom and culture. *REiLA: Journal of Research and Innovation in Language, 3*(1), 42–51. https://doi.org/10.31849/reila.v3i1.5091

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). Sage Publications.

Lamido, A. (2014). *Ya'yan safara*. Sauki Press.

Mukhtar, I. (2024). *An introductory Hausa linguistics*. Bayero University Press.

Newman, P. (2000). *The Hausa language: An encyclopedic reference grammar*. Yale University Press.

Sa'id, B. (2019). *Ka'idojin rubutun Hausa: Jagora ga marubuta*. Benchmark Publishers.

Schiffrin, D. (1987). *Discourse markers*. Cambridge University Press.

Shuaibu, A. (2025). Hausa language on Facebook: A linguistic analysis of grammatical errors and their effects on orthography. *Journal of Second and Multiple Language Acquisition-JSMULA, 13*(1), 730–749. https://doi.org/10.5281/zenodo.15570333

Yahaya, I. Y. (1988). *Nahawun Hausa*. University Press Limited.

Zając, P. (2024). Pragmatic marker *to* in Hausa (West Chadic, A.1; Nigeria). *LingBaW: Linguistics Beyond and Within, 10*, 247–263. https://doi.org/10.31743/lingbaw.18022

EDITOR-IN-CHIEF
Prof. S. T. Addo [GHANA]