

Nigerian Pidgin English: Grammar and Core Vocabulary with Script Representation

Flyxion

?today?

Abstract

Nigerian Pidgin English is a major contact language spoken by tens of millions of people across Nigeria and neighboring regions of West Africa. Although frequently described in popular discourse as a “pidgin” in the sense of a simplified or auxiliary code, it exhibits systematic phonology, stable morphosyntax, productive aspect marking, and discourse strategies comparable to other creole and post-creole continua. Its grammatical organization cannot be reduced to defective English; rather, it reflects a structured outcome of sustained multilingual contact.

This essay offers a formal introduction to Nigerian Pidgin as a rule-governed linguistic system. It surveys its segmental phonology, pronominal paradigm, tense–aspect system, serial verb constructions, negation patterns, interrogative formation, and high-frequency lexical domains. In addition to descriptive analysis, the study develops parallel representational systems in both Latin phonemic transcription and a constrained Arabic-script transliteration framework. The goal is not conversational fluency but structural orientation: to demonstrate that Nigerian Pidgin possesses internal regularities and distributional constraints sufficient to sustain grammatical description, cross-script encoding, and computational modeling.

1 Introduction

Languages that arise from contact are often mischaracterized as transitional or structurally incomplete. Such descriptions obscure the fact that contact languages stabilize through regularization, constraint, and functional pressure. Once stabilized, they display patterned morphosyntax and systematic phonology no less than historically older literary languages. Nigerian Pidgin is exemplary in this regard.

The purpose of this study is not to rehearse sociolinguistic debates over classification, nor to provide a phrasebook-style inventory of expressions. Instead, it adopts a structural perspective. Nigerian Pidgin is treated here as a linguistic system whose components can be described in terms of phoneme inventories, morphosyntactic categories, and combinatorial rules. Its apparent transparency to speakers of English masks underlying reorganizations of tense–aspect marking, argument structure, and discourse management that are neither accidental nor ad hoc.

The approach taken in the following sections proceeds from foundational description to grammatical architecture. Phonological inventory establishes the sound system. Pronominal forms and aspectual markers reveal core syntactic organization. Serial verb constructions illustrate clause-level compositionality. Negation and interrogative formation demonstrate rule-governed transformations rather than lexical substitution. Vocabulary is presented not as a memorization list but as evidence of semantic structuring within recurrent domains.

By foregrounding structural regularity, this introduction aims to reposition Nigerian Pidgin within linguistic analysis as a coherent grammar rather than a derivative code. Such treatment is essential not only for descriptive accuracy but also for orthographic and computational extension. Accordingly, the study develops parallel Latin phonemic representations and a formally constrained Arabic-script transliteration system. These representational layers do not precede structure; they depend upon it. The primary object of study remains the internal organization that makes systematic cross-script encoding possible.

2 Historical and Linguistic Background

Nigerian Pidgin emerged from sustained contact between English speakers and diverse West African language communities beginning in the early colonial period. It developed as a trade and interethnic communication language and gradually stabilized across regions. While

English supplies much of the lexical material, the grammar reflects substrate influence from Niger–Congo languages, especially in aspect marking, serial verb constructions, and pronoun systems.

Today Nigerian Pidgin functions as a lingua franca in urban centers such as Lagos, Port Harcourt, Warri, and Benin City. It is used in music, radio broadcasting, informal media, comedy, and everyday interaction. Although it does not yet possess full official status in education, its sociolinguistic vitality is undeniable.

Crucially, Nigerian Pidgin is not “broken English.” It is a rule-governed system. Many of its features represent simplification relative to Standard English morphology, but simplification does not imply absence of structure. Instead, complexity is reorganized, particularly in the tense–aspect system.

3 Phonological Tendencies

Nigerian Pidgin pronunciation reflects both English input and substrate phonological systems. Certain consonant clusters are simplified. Word-final consonants may be reduced. Vowel quality is more stable and less diphthongized than in many English dialects.

Examples:

school → *skul*

friend → *fren*

thing → *tin*

The interdental fricatives /θ/ and /ð/ are typically replaced by /t/ or /d/ respectively:

three → *tree*

that → *dat*

Stress patterns are more even, and rhythm often reflects syllable-timed tendencies common in West African languages rather than the stress-timed rhythm typical of British English.

4 Core Morphosyntactic Structure

4.1 Basic Word Order

Nigerian Pidgin exhibits canonical Subject–Verb–Object (SVO) word order, consistent with English surface structure but lacking the inflectional morphology characteristic of Standard English.

I chop rice.
(I eat rice.)

Verbs do not inflect for person or number. There is no subject–verb agreement morphology:

I go.
You go.
Dem go.

The absence of inflection does not reduce grammatical clarity because temporal and aspectual distinctions are encoded analytically through preverbal particles.

4.2 Pronoun System

The pronominal system is stable and invariant across most dialectal varieties:

Singular	Plural
I	We
You	Una
E (he/she/it)	Dem

Notable features include:

- The absence of gender distinction in the third person singular: *e* refers to he, she, or it.
- The second person plural form *una*, which has no equivalent in Standard English.
- The third person plural *dem*, which may function both as pronoun and plural marker.

Example:

Dem dey come.
(They are coming.)

Plural marking may also appear post-nominally:

Di boy dem.
(The boys.)

Here *dem* functions as a pluralizer rather than a pronoun.

4.3 Copula and Existential Constructions

The copular system differs from English. The verb *be* does not function as an inflected auxiliary. Instead, several strategies appear.

Zero copula:

I happy.
(I am happy.)

Na-cleft structure:

Na doctor e be.
(It is a doctor.)

Na serves as a focus or identificational marker.

Existential construction:

Get problem.
(There is a problem.)

or

E get problem.

4.4 Verb Invariance

Verbs in Nigerian Pidgin do not inflect for tense or agreement. Instead, temporal reference is either contextually inferred or expressed through preverbal markers, discussed in the next section.

Examples:

I chop yesterday.
(I ate yesterday.)

I go tomorrow.

(I will go tomorrow.)

Temporal adverbs frequently carry the burden of time specification in the absence of morphological tense.

5 Tense Aspect and Mood

Nigerian Pidgin relies primarily on preverbal particles rather than inflectional morphology to express tense, aspect, and modality. These particles occur between the subject and the verb and remain invariant across persons.

5.1 Progressive Aspect: *dey*

The particle *dey* marks progressive or ongoing action.

I dey chop.

(I am eating.)

Dem dey work.

(They are working.)

The progressive is thus analytic rather than inflectional. The verb itself remains unmodified.

5.2 Past Marker: *bin*

The particle *bin* marks anterior or past reference.

I bin go.

(I went / I had gone.)

Dem bin see am.

(They saw him/her/it.)

In many contexts, past time may also be inferred from adverbials without overt marking. The use of *bin* tends to emphasize temporal distance or narrative sequencing.

5.3 Future Marker: *go*

The particle *go* expresses future reference or intention.

I go come tomorrow.
(I will come tomorrow.)

E go rain.
(It will rain.)

Unlike English “will,” *go* does not inflect and does not require auxiliary inversion in questions.

5.4 Completive Aspect: *don*

The particle *don* marks completed action, roughly corresponding to the English present perfect or perfective.

I don chop.
(I have eaten / I already ate.)

Dem don finish.
(They have finished.)

Don emphasizes completion rather than temporal location.

5.5 Habitual Aspect

Habitual meaning may be expressed through context or through repeated use of *dey* in certain dialects:

I dey go market every day.
(I go to the market every day.)

The distinction between progressive and habitual is often context-driven rather than morphologically encoded.

5.6 Modal Constructions

Modal meanings are typically expressed lexically rather than through auxiliary inflection.

Ability:

I fit do am.
(I can do it.)

Obligation:

You for go.
(You should have gone / You ought to go.)

Necessity:

You must go.
(You must go.)

The system is thus largely analytic. Grammatical categories are expressed through separate lexical items rather than bound morphology.

5.7 Particle Ordering

Particles may combine in structured sequences. For example:

I bin dey work.
(I was working.)

Dem don dey wait.
(They have started waiting / They are already waiting.)

The ordering reflects temporal scope, with past or completive markers typically preceding progressive marking.

6 Negation and Interrogative Structure

6.1 Negation

Negation in Nigerian Pidgin is typically expressed with the preverbal particle *no*.

I no go.
(I did not go / I do not go.)

Dem no dey work.
(They are not working.)

Negation precedes aspectual markers when both are present:

I no bin see am.

(I did not see him/her/it.)

The structure remains analytic and invariant; the verb itself does not change form.

6.2 Negative Imperatives

Prohibition is typically formed using *no* or *make no*:

No go there.

(Don't go there.)

Make you no talk.

(Do not speak.)

6.3 Yes–No Questions

Yes–no questions are generally formed through intonation rather than auxiliary inversion.

You dey come?

(Are you coming?)

Dem don finish?

(Have they finished?)

There is no inversion comparable to English “Are you coming?” The surface structure remains declarative, and interrogative force is conveyed prosodically.

6.4 Wh-Questions

Interrogative words are typically fronted but do not trigger subject–auxiliary inversion.

Wetyn	What
Who	Who
Where	Where
When	When
Why	Why
How	How

Examples:

Wetyn you dey do?
(What are you doing?)

Where dem go?
(Where did they go?)

Again, the verb remains uninflected.

6.5 Focus and Emphasis with *na*

The particle *na* serves as a focus marker and clefting device.

Na me do am.
(It was I who did it.)

Na today we go start.
(It is today that we will begin.)

The structure resembles cleft constructions in other creole languages and in certain Niger–Congo substrate languages.

7 Serial Verb Constructions

Serial verb constructions are common and reflect substrate influence from West African languages. Multiple verbs may appear in sequence without conjunctions, expressing closely linked actions.

I go buy book.
(I will go and buy a book.)

Dem carry come.
(They brought it.)

I take knife cut am.
(I used a knife to cut it.)

In the final example, *take* functions as an instrumental marker rather than a fully independent verb in the English sense.

Serial constructions allow compact expression of direction, instrumentality, causation, and sequential action without subordination or conjunction.

8 Core Vocabulary and Lexical Domains

Although much of Nigerian Pidgin vocabulary derives historically from English, lexical meaning often diverges in scope, usage, or pragmatic force. The lexicon also incorporates substrate and regional elements.

8.1 Basic Verbs

A small set of high-frequency verbs carries much of the communicative load:

<i>chop</i>	eat
<i>go</i>	go
<i>come</i>	come
<i>see</i>	see
<i>hear</i>	hear
<i>talk</i>	speak
<i>carry</i>	carry / bring
<i>take</i>	take / use
<i>give</i>	give
<i>make</i>	make / cause

Many verbs have broader semantic range than their English equivalents. For example, *chop* may also mean “consume” metaphorically:

Dem chop money.

(They embezzled money.)

8.2 Common Nouns

<i>pikin</i>	child
<i>woman</i>	woman
<i>man</i>	man
<i>house</i>	house
<i>money</i>	money
<i>wahala</i>	trouble
<i>palava</i>	problem / dispute

Wahala and *palava* are widely used to indicate difficulty or complication.

8.3 Function Words and Particles

Function words play a critical structural role.

<i>dey</i>	progressive marker
<i>don</i>	completive
<i>bin</i>	past
<i>go</i>	future
<i>no</i>	negation
<i>na</i>	focus marker
<i>wey</i>	relative marker

Relative clauses are typically formed using *wey*:

Di man wey come yesterday.
(The man who came yesterday.)

8.4 Discourse Markers

Nigerian Pidgin includes a range of discourse markers used for emphasis, alignment, and pragmatic nuance:

<i>sha</i>	mild emphasis / concession
<i>o</i>	emphasis or emotional marking
<i>abeg</i>	please / request softener
<i>now</i>	intensifier or focus particle

Examples:

I go come sha.
(I will come, anyway.)

Abeg help me.
(Please help me.)

Na today now.
(It is today indeed.)

These markers operate pragmatically rather than syntactically, contributing to tone and stance.

9 Semantic Extension and Pragmatic Usage

Lexical meaning in Nigerian Pidgin frequently extends beyond its etymological English base. Words may broaden, narrow, or shift in pragmatic value.

For example:

How far?
(Literal: How far?)
Meaning: How are things? What is happening?

You dey?
(Are you present? / Are you available?)

Meaning often depends on context, shared cultural understanding, and prosodic cues rather than morphological marking.

10 Grammatical Enrichment

10.1 Reduplication

Reduplication functions as an intensifier, distributive marker, or expressive device. It is common in many West African languages and appears productively in Nigerian Pidgin.

Intensification:

small small
(little by little)

fine fine
(very fine / extremely beautiful)

Distributive meaning:

Dem come one one.
(They came one by one.)

Reduplication adds semantic nuance without inflectional morphology.

10.2 Comparative and Superlative Structures

Comparison does not rely on inflectional endings such as “-er” or “-est.” Instead, comparison is expressed analytically.

Comparative with *pass*:

Dis one big pass dat one.
(This one is bigger than that one.)

Pass functions as a comparative marker meaning “more than.”

Superlative with *pass all*:

Na im fine pass all.
(He/she is the most beautiful.)

Comparison is thus relational rather than morphologically marked.

10.3 Quantifiers and Plurality

Plural marking is optional and often contextually determined. The plural marker *dem* may follow a noun.

Book dem dey table.
(The books are on the table.)

Quantifiers include:

<i>plenty</i>	many / much
<i>small</i>	little
<i>all</i>	all
<i>every</i>	every

Examples:

Plenty people come.
(Many people came.)

Small water remain.
(A little water remains.)

10.4 Possession

Possession is typically expressed through juxtaposition rather than apostrophe marking.

Di man house.
(The man's house.)

My papa car.
(My father's car.)

The possessive relation is interpreted structurally rather than morphologically.

10.5 Relative Clauses

The relative marker *wey* introduces subordinate descriptive clauses.

Di woman wey dey talk.
(The woman who is speaking.)

No relative pronoun inflection occurs. The marker remains invariant.

10.6 Conditional Constructions

Conditional statements are typically formed with *if* or with serial structures.

If you come, I go happy.
(If you come, I will be happy.)

Hypothetical or counterfactual nuance may be conveyed through context or through use of *for*.

You for tell me.

(You should have told me.)

10.7 Imperatives and Causatives

Imperatives are generally bare verbs:

Come here.

Carry am.

Causation frequently uses *make*:

Make I go.

(Let me go.)

Make dem start.

(Let them begin.)

Here *make* functions as a causative or permissive marker.

11 Pragmatics and Politeness Strategies

Nigerian Pidgin operates within a rich pragmatic ecology in which politeness, hierarchy, solidarity, and emotional stance are frequently conveyed through particles, repetition, tone, and lexical choice rather than through formal grammatical marking.

11.1 Softening Devices

The particle *abeg* functions as a politeness marker, mitigating requests and commands.

Abeg give me water.

(Please give me water.)

It can also appear mid-sentence to soften disagreement:

Abeg, no vex.

(Please, do not be angry.)

11.2 Emphatic and Affective Particles

The sentence-final particle *o* marks emotional emphasis, insistence, or affective engagement.

Come here o.

(Come here, I insist / urgently.)

The particle *sha* may signal concession or mild contrast:

I go try sha.

(I will try anyway.)

11.3 Solidarity and Alignment

Greetings and discourse openings often establish social alignment.

How far?

(How are things?)

You dey?

(Are you there? / Are you around?)

Such expressions function less as literal questions and more as relational markers.

11.4 Respect and Hierarchy

Honorific marking is generally lexical rather than morphological. Terms such as *oga* (boss) or *madam* index respect.

Oga, I don finish.

(Boss, I have finished.)

Politeness strategies rely heavily on intonation, context, and relational expectations.

12 Code-Switching and English Overlap

Nigerian Pidgin exists in a dynamic continuum with Standard Nigerian English and with regional languages such as Yoruba, Igbo, and Hausa. Speakers frequently alternate between varieties depending on context, audience, and social positioning.

12.1 Lexical Mixing

Code-switching may involve insertion of English technical vocabulary into Pidgin structure.

I go submit di assignment tomorrow.

The syntactic frame remains Pidgin while lexical items derive from Standard English.

12.2 Register Shifting

In formal contexts, speakers may shift toward Standard English morphology while retaining Pidgin rhythm or particles. In informal contexts, greater reliance on Pidgin structures appears.

This fluidity complicates attempts at strict grammatical boundary drawing. Nigerian Pidgin functions not as an isolated code but as part of a sociolinguistic spectrum.

13 Regional Variation

Although broadly intelligible across Nigeria, Nigerian Pidgin exhibits regional phonological and lexical variation.

In southern urban centers such as Warri, Port Harcourt, and Lagos, Pidgin often functions as a primary lingua franca. In northern regions, usage may be more restricted and influenced by Hausa structures.

Lexical items such as *wahala* may carry slightly different pragmatic weight across regions. Pronunciation and intonation patterns vary significantly.

The language is thus internally heterogeneous rather than uniform.

14 Orthography and Standardization

Nigerian Pidgin lacks a fully standardized orthography. Spellings vary according to phonetic intuition and influence from English conventions.

For example, the progressive marker may appear as:

dey, de, or day

Similarly, *wetin* may appear as *wetyn*.

Despite this variation, there are emerging efforts toward standardization in media, literature, and education. Broadcast journalism and online platforms increasingly employ consistent spelling conventions.

Orthographic fluidity reflects the primarily oral historical transmission of the language.

15 Annotated Sample Dialogue

A: *How far? You dey?*

B: *I dey. Wetin happen?*

A: *I bin call you yesterday but you no pick.*

B: *Ah, I bin dey work. I don finish now.*

A: *Good. Make we go market small small.*

B: *Okay. I go come now.*

Analysis:

How far? functions as greeting rather than literal distance inquiry.

I dey marks presence or availability.

bin marks past action.

no marks negation.

don marks completion.

make we forms a hortative construction.

go marks future intention.

16 Phonology and an Alternative Orthographic Proposal

16.1 Phonological Overview

Nigerian Pidgin phonology reflects substrate influence from West African languages and superstrate influence from English. The sound system is generally simpler in syllable structure than Standard English, with reduced consonant clusters and strong preference for open syllables in many regional varieties.

Consonant clusters at word-final position are frequently simplified:

ask → *aks*

child → *chail*

Vowel quality tends toward a reduced inventory compared to English, and diphthongs are often monophthongized depending on speaker background.

Tone is not lexically contrastive in the way it is in Yoruba or Igbo, but intonation carries significant pragmatic meaning, particularly in interrogatives and emphasis.

16.2 Syllable Structure

The dominant syllable structure approximates (C)V(C), though final consonants may be weakened or deleted in casual speech. Many borrowings from English undergo phonotactic adaptation to align with West African patterns.

Reduplication and particle-based grammar further reinforce rhythmic and prosodic regularity.

16.3 The Case for Orthographic Reconsideration

The current Latin-based orthography inherits English spelling conventions, which introduce inconsistency between sound and symbol. Variants such as *dey*, *de*, and *day* illustrate instability.

An orthography more closely aligned to phonetic realization could reduce ambiguity. One possible direction is transliteration into Arabic script (Ajami-style adaptation), which has historical precedent in West Africa across Hausa and other languages.

Such a system would not imply cultural replacement but phonetic alignment. Arabic script represents consonantal structure efficiently and can be modified with diacritics to capture vowel distinctions.

16.4 Proposed Arabic-Based Transliteration Framework

The following is a preliminary mapping based on approximate phonetic correspondence.

Consonants:

b	ب
d	د
f	ف
g	غ or گ (modified)
k	ك
l	ل
m	م
n	ن
p	پ (Persian extension)
r	ر
s	س
t	ت
w	و
y	ي

Vowels would be represented through short vowel diacritics and long vowel letters:

a	َ
i	ِ
u	ُ
long a	ا
long i	ي
long u	و

16.5 Illustrative Examples

I dey could be rendered:

ادي

Wetin happen?

وتن هپن

I no go

انوغو

This transliteration prioritizes phonetic transparency over etymological fidelity.

16.6 Advantages and Challenges

Advantages include closer sound-to-symbol alignment and potential integration into broader Ajami literacy traditions in West Africa.

Challenges include:

- Dialectal vowel variation,
- Absence of standardized tone marking,
- Sociopolitical sensitivity surrounding script change,
- Technical encoding and digital adoption.

The proposal remains exploratory. It seeks phonological coherence rather than cultural substitution.

16.7 Theoretical Implication

Orthography shapes perception of linguistic structure. A phonetic Arabic-based script would foreground syllabic rhythm and particle-based grammar rather than English lexical ancestry. It would visually decouple Nigerian Pidgin from Standard English, potentially reinforcing its autonomy as a linguistic system.

17 Toward a Phonetic Ajami Orthography for Nigerian Pidgin

17.1 Design Principles

The proposed Arabic-based transliteration system is not etymological but phonetic. Its primary objective is one-to-one mapping between sound and symbol, thereby minimizing the orthographic opacity inherited from English.

Three structural principles guide the system:

First, consonants are mapped according to phonetic similarity rather than English spelling convention.

Second, vowels are represented consistently through diacritics and long-vowel letters, prioritizing phonemic transparency.

Third, grammatical particles are preserved visually as independent units, reinforcing the analytic structure of Nigerian Pidgin.

17.2 Consonant Mapping

The consonant inventory largely overlaps with Arabic but requires extended letters for sounds not native to Classical Arabic. Persian extensions such as پ for /p/ are employed where necessary to preserve phonemic contrast.

For example:

b	ب
d	د
f	ف
g	گ / غ
k	ك
l	ل
m	م
n	ن
p	پ
r	ر
s	س
t	ت
w	و
y	ي
h	ه

Affricates such as /ch/ and /j/ may be represented using adapted forms, for example تش and ج, respectively.

17.3 Vowel Representation

Short vowels are indicated through diacritics:

a = i = u =

Long vowels are represented through ا, ي, and و.

For example:

آي دي
(I dey)

Here آي captures the diphthongal realization of “I,” while دي captures the progressive particle /de /.

17.4 Grammatical Particles in Script

A significant structural advantage of this orthography is the visual stabilization of analytic particles.

نَا سُو
(Na so)

دِم دُونْ كَم
(Dem don come)

Particles such as نَا (na), دُونْ (don), and دي (dey) become visually discrete units, reinforcing their grammatical independence.

This representation visually encodes the analytic structure of tense and aspect marking more transparently than English-based spelling.

17.5 Phonotactic Adaptation

The orthographic system resolves cluster reduction by reflecting actual spoken realization rather than English orthography.

For example:

وَيْتِنْ
(Wetin)

This avoids misleading English silent letters and aligns orthography more closely with speech rhythm.

17.6 Semantic Autonomy Through Script

By removing English orthographic anchoring, the script visually decouples Nigerian Pidgin from Standard English. This has structural consequences.

The language appears as a system in its own right rather than a deviation from English. Particles gain structural salience. Prosodic rhythm becomes more apparent.

Orthographic reform therefore does not merely change script; it reorganizes perception of grammatical structure.

17.7 Limitations and Technical Questions

Dialectal vowel variation presents challenges. The distinction between /e/ and / / or between /o/ and / / may require additional diacritic marking if phonemic contrast is deemed significant.

Digital encoding requires standardization of extended letters such as and across platforms.

Sociolinguistic reception would likely vary across regions and communities.

17.8 Algorithmic Transliteration

The system is suitable for deterministic transliteration because it operates primarily on phoneme-to-grapheme mapping rather than etymological mapping.

A bidirectional algorithm would require:

A phonemic tokenizer for Latin Pidgin input.

A rule-based consonant mapping.

A vowel diacritic insertion rule sensitive to syllable structure.

Optional tone marking if future refinement demands it.

The resulting system is structurally simpler than English orthography because it does not encode historical spelling residue.

18 Phoneme Inventory and IPA Mapping

This section proposes a working phoneme inventory for Nigerian Pidgin suitable for orthographic design and then specifies an IPA-to-script mapping consistent with the transliteration practice illustrated in the corpus above. The goal is not to legislate a single “true” pronunciation for all regions, but to provide a coherent baseline inventory that supports deterministic transliteration.

18.1 Working Phoneme Inventory

Nigerian Pidgin exhibits regional variation, but a practical orthography can be built around a core consonant system and a modest vowel system. The inventory below is intended as an operational compromise: fine-grained enough to support stable transcription, coarse-grained enough to remain dialect-robust.

Consonants

/p, b, t, d, k, g, m, n, ŋ, f, v, s, z, h, w, j, l, r, t, d/

Interdental fricatives /θ/ and /ð/ are excluded, since they are not phonemic in most Nigerian Pidgin varieties and are typically realized as alveolar stops, as reflected in forms such as *dat* and *tree*.

Vowels A five-vowel system provides stable coverage:

/i, e, a, o, u/

A seven-vowel refinement with // and // may be optionally encoded, but orthographic stability does not require it. Diphthongs are treated as vowel sequences, with /ai/ especially frequent.

18.1.1 Latin Baseline (Phonemic)

Before transliteration into Arabic script, phonemic forms are normalized into a narrow Latin baseline that avoids English spelling conventions. This baseline stabilizes input and prevents interference from English orthography.

IPA	Latin	Example
/p/	p	pikin
/b/	b	boy
/t/	t	tin
/d/	d	dey
/k/	k	ko
/g/	g	go
/t /	ch	chop
/d /	j	enjoy
/f/	f	fit
/v/	v	vex
/s/	s	so
/z/	z	zaa (dialectal)
/h/	h	how
/m/	m	make
/n/	n	no
/ŋ/	ng	sing
/l/	l	lie
/r/	r	far
/w/	w	wetin
/j/	y	you

This level is purely phonemic. English spellings such as *know*, *three*, or *that* are excluded in favor of *no*, *tri*, and *dat*.

18.1.2 Arabic Script Mapping

Arabic transliteration is derived mechanically from the Latin baseline. Only standard Arabic letters are used.

Consonants

IPA	Script	Note
/p/	ب	approximated as /b/
/b/	ب	direct
/t/	ت	direct
/d/	د	direct
/k/	ك	direct
/g/	ج	phonetic approximation
/f/	ف	direct
/v/	ف	merged with /f/
/s/	س	direct
/z/	ز	direct
/h/	ه	direct
/m/	م	direct
/n/	ن	direct
/ŋ/	ننج	cluster representation
/l/	ل	direct
/r/	ر	direct
/w/	و	direct
/j/	ي	direct
/t /	تش	digraph
/d /	ج	direct

No additional Persian letters are introduced. All approximations are handled internally through clustering or phonemic merger.

Vowels Short vowels are represented by diacritics when phonological clarity requires explicit marking. The vowel /a/ is written with [َ] (fatha), /i/ and /e/ with [ِ] (kasra), and /u/ and /o/ with [ُ] (damma).

In unstressed environments or highly predictable contexts, short vowels may be omitted in fluent writing; however, full marking ensures deterministic back-mapping to the Latin baseline and eliminates ambiguity in pedagogical contexts.

Long vowels are represented by the corresponding matres lectionis. The vowel /a / is written with ا, /i / with ي, and /u / with و. Where contrastive vowel length is not phonemic, these

long-vowel letters function primarily as quality markers rather than as duration markers.

Diphthongs are written as ordered sequences of vowel letters. The diphthong /ai/ is represented as آي, while /au/ is written as آو. In environments where a glottal onset is phonetically salient, initial آ provides a stable orthographic anchor.

18.1.3 Worked Derivations

The full transliteration pipeline proceeds from surface form to phonological representation, from phonology to Latin baseline, and finally from the Latin baseline to Arabic script.

Progressive The form *I dey* is analyzed phonologically as /ai de/, normalized to the Latin baseline *ai de*, and rendered orthographically as آي دي.

Negation The expression *I no know* corresponds phonologically to /ai no no/, normalizes to *ai no no*, and is written as آي نو نو.

Wh-question The form *Wetin dey happen* is analyzed as /wetin de apen/, normalized to *wetin de apen*, and rendered as ويتن دي ابن.

Consonant clusters are either resolved through vowel insertion consistent with attested pronunciation or indicated with sukūn, as in , where pedagogically useful to preserve segmental transparency.

18.1.4 Functional Stability

High-frequency grammatical particles maintain fixed orthographic forms. The progressive marker *dey* is written as دي, the completive marker *don* as دون, the copular or focus particle *na* as نا, and the negator *no* as نو.

Stability at this level reinforces morphosyntactic transparency and ensures that analytic tense and aspect marking remains visually distinct.

Stability at this level reinforces morphosyntactic transparency and reduces ambiguity in analytic tense and aspect marking.

18.1.5 Orthographic Principle

The system is strictly phonological. Each grapheme corresponds to a surface phoneme in the working inventory. English spelling conventions are excluded from derivation. Dialectal variation can be layered through optional diacritics without altering the core mapping.

The result is a closed, internally consistent transliteration framework: inventory \rightarrow Latin phonemic normalization \rightarrow Arabic graphemic realization. Each stage precedes the next.

18.2 Deterministic Transliteration Rules

This subsection specifies a deterministic procedure for transliterating a restricted Latin baseline into an Arabic-only orthography. The procedure is designed to be implementable as a finite sequence of rewrite rules. It assumes that the Latin input is already normalized to a phonetic baseline rather than English spelling.

18.2.1 Input Normalization

Let the input string be a sequence of tokens separated by spaces. Each token is assumed to be lower-case Latin characters drawn from the set

$$\{a, e, i, o, u, b, d, f, g, h, j, k, l, m, n, r, s, t, w, y\}$$

together with the digraphs sh and ch and apostrophe-free orthography. If the source uses English-residue spellings, these must be normalized first.

A minimal normalization map that removes common English residue is:

$$\text{th} \mapsto \text{t}, \quad \text{ph} \mapsto \text{f}, \quad \text{ck} \mapsto \text{k}, \quad \text{qu} \mapsto \text{kw}.$$

This stage is not a model of pronunciation; it is a structural stabilizer for transliteration.

18.2.2 Tokenization

Let the token sequence be T_1, \dots, T_N . Transliteration applies tokenwise. Punctuation may be retained and copied directly after the corresponding token.

18.2.3 Consonant Mapping

Within each token, consonantal segments are mapped deterministically by the following substitution table.

Latin	Arabic	Comment
b	ب	
d	د	
f	ف	
g	ج	Arabic-only approximation
h	ه	
j	ج	shares glyph with g; resolved by context
k	ك	
l	ل	
m	م	
n	ن	
r	ر	
s	س	
t	ت	
w	و	consonantal /w/
y	ي	consonantal /y/
sh	ش	
ch	تش	Arabic-only cluster encoding

The map is applied greedily left-to-right with longest-match priority, so sh and ch are rewritten before single-letter consonants.

18.2.4 Vowel Strategy

Two regimes are possible. The first is a readability-first regime that writes vowels as letters where possible. The second is a precision-first regime that uses diacritics systematically. The transliteration system in this essay adopts the readability-first regime as default, with diacritics treated as optional refinement.

The readability-first vowel map is:

Latin	Arabic	Comment
a	ا	default long
i	ي	default long
u	و	default long
e	ي	nearest-grapheme approximation
o	و	nearest-grapheme approximation

This strategy produces a stable orthography without requiring diacritic entry. It intentionally sacrifices fine vowel quality distinctions, which in Nigerian Pidgin are frequently predictable from lexical and syntactic context.

An optional refinement permits systematic short-vowel marking in pedagogical mode. Under this refinement, the vowel /a/ is marked with [˘], the vowels /i/ and /e/ with [˙], and the vowels /u/ and /o/ with [˘]. Such marking may be employed when the writer wishes to avoid ambiguity or to ensure fully deterministic phonological recovery.

18.2.5 Special Lexical Tokens

Certain high-frequency function words are best treated lexically in order to preserve recognizable forms and avoid overgeneration. Define a small lexical override dictionary \mathcal{D} , applied prior to general rules:

Latin	Arabic
i	آي
na	نا
no	نو
dey	دي
don	دون
dem	ديم
wetin	ويتين
abeg	ابج

This procedure ensures that analytic particles remain visually discrete, reinforcing their grammatical independence and supporting the pedagogical function of the script.

18.2.6 Ambiguity Management

Because Arabic-only constraints collapse certain contrasts, two systematic ambiguities arise.

The first is the collapse of /g/ and /j/ under ج. Resolution is lexical: in practice, Nigerian Pidgin has relatively few minimal pairs dependent on this contrast, and the intended reading is recoverable from context. If disambiguation is required, one may optionally reserve غ for /g/ and ج for /j/ as an internal convention, while still remaining within Arabic letters.

The second is the collapse of /p/ into ب if the Latin baseline includes p. Under Arabic-only constraints, p is rewritten as ب. This is historically normal in many Ajami systems. If one wishes to preserve /p/ without non-Arabic letters, the only option is a diacritic convention on ب, but this is best treated as a later refinement rather than a core requirement.

18.2.7 Algorithmic Order

The transliteration pipeline is therefore:

First, apply lexical overrides \mathcal{D} tokenwise. Second, apply normalization to remove English residue patterns. Third, apply greedy consonant mapping with longest-match priority. Fourth, apply the vowel strategy, either readability-first or pedagogical-diacritic mode. Finally, reattach punctuation and preserve spacing.

This order ensures that high-frequency particles stabilize early, consonantal structure is preserved deterministically, and vowels are added in a predictable manner.

18.2.8 Worked Derivations

Example 1: *i dey* Application of the lexical override dictionary yields *i* → آي and *dey* → دي. The resulting form is therefore آي دي.

Example 2: *wetin dey happen* Lexical override first yields *wetin* → ويتين and *dey* → دي. The remaining token *happen* is processed by consonant mapping, where h corresponds to ه, p to ب, and n to ن. Under the readability-first vowel strategy, the stable orthographic output is هابن. The complete expression is therefore rendered as ويتين دي هابن؟.

Example 3: *dem don come* Lexical override yields *dem* → ديم and *don* → دون. The remaining token *come* undergoes consonant mapping, where c corresponds to ك and m to م. Vowel mapping inserts و for /o/ and ي for /e/, producing the intermediate form كومي. A readability adjustment may optionally omit the final ي to reflect common pronunciation, yielding كوم. The full expression is thus written as ديم دون كوم.

The derivations illustrate that the system is deterministic at the rule level while still permitting a controlled layer of phonological smoothing for high-frequency words.

19 Comparison with Hausa Ajami

Any proposal for an Arabic-based orthography for Nigerian Pidgin necessarily invites comparison with Hausa Ajami, the long-standing Arabic-script tradition used to write Hausa across northern Nigeria and the wider Sahel. The comparison is instructive because Hausa Ajami represents one of the most historically successful adaptations of Arabic script to a non-Arabic phonological system in West Africa.

19.1 Historical Orientation

Hausa Ajami emerged within Islamic scholarly networks and was used for poetry, legal documents, correspondence, and religious instruction. It developed organically over centuries, rather than through centralized planning. Orthographic conventions stabilized gradually through repeated manuscript transmission.

In contrast, the proposed Pidgin Ajami system is explicitly planned. It begins with a phonological inventory and a deterministic transliteration algorithm. The Hausa system evolved bottom-up; the Pidgin system is top-down.

19.2 Phonological Adaptation

Hausa contains phonemic contrasts not present in Arabic, including ejectives and implosives, as well as distinct /p/ and /g/ sounds. Hausa Ajami resolved these gaps in several ways.

First, some traditions introduced modified letters through dotting conventions.

Second, some contrasts were collapsed orthographically and disambiguated contextually.

Third, vowel length distinctions were often marked more carefully than in unvowelled Arabic.

The proposed Pidgin Ajami system aligns most closely with the second strategy. Under Arabic-only constraints, /p/ merges with ب and /g/ may be represented by ج or غ. As in Hausa Ajami, phonological recovery is expected to occur through lexical familiarity rather than perfect graphemic fidelity.

The structural similarity lies in functional tolerance for controlled ambiguity.

19.3 Vowel Representation

Hausa distinguishes long and short vowels phonemically. In many Ajami manuscripts, long vowels are written explicitly with ا, ي, and و, while short vowels may be omitted or marked with diacritics depending on genre.

Nigerian Pidgin does not rely heavily on vowel length contrasts for lexical differentiation. The proposed Pidgin Ajami therefore adopts a readability-first vowel strategy, writing core vowels explicitly while permitting optional diacritics for pedagogical precision.

Both systems therefore prioritize functional literacy over phonetic exhaustiveness.

19.4 Morphological Transparency

Hausa is morphologically richer than Nigerian Pidgin. It contains noun class remnants, plural patterns, and verbal derivations. Hausa Ajami thus encodes morphological alternations within stems.

Nigerian Pidgin, by contrast, is analytically structured. Tense, aspect, and modality are particle-based rather than inflectional. This produces a different orthographic opportunity.

In Pidgin Ajami, particles such as نا (na), دي (dey), دون (don), and نو (no) become visually stable units. Because these particles are invariant, the script reinforces analytic grammar rather than morphological alternation.

Hausa Ajami must accommodate internal stem alternation. Pidgin Ajami highlights syntactic particles.

19.5 Sociolinguistic Position

Hausa Ajami historically functioned alongside Arabic literacy and later alongside Latin-script Hausa. It did not replace Latin orthography but coexisted with it.

A Pidgin Ajami system would likely occupy a similar sociolinguistic niche: supplemental rather than replacement orthography. It could function in artistic, pedagogical, or cultural contexts without displacing established Latin usage.

The comparison therefore suggests that script plurality is not structurally destabilizing. West African linguistic ecology already supports dual-script traditions.

19.6 Structural Divergence

Despite parallels, there is a critical difference.

Hausa is an indigenous Niger–Congo language with a long-standing literary identity. Ajami writing reinforced an existing linguistic core.

Nigerian Pidgin is a contact language with English lexical ancestry and multi-ethnic substrate influence. Writing it in Arabic script visually decouples it from English orthography. This is not merely phonetic adaptation; it is perceptual reclassification.

Hausa Ajami preserves Hausa as Hausa. Pidgin Ajami repositions Pidgin as structurally autonomous.

19.7 Implications for Standardization

The Hausa case demonstrates that:

Orthographic conventions stabilize through repeated communal usage rather than abstract design alone.

Phoneme–grapheme mismatch does not prevent literacy if functional load remains manageable.

Script change does not erase Latin literacy but can coexist with it.

For Nigerian Pidgin, the lesson is not that a perfect mapping is required. Rather, consistency and repeatability are more important than exhaustiveness. A deterministic transliteration framework, once adopted in sustained textual practice, would likely converge toward stable norms just as Hausa Ajami did historically.

20 Deterministic Reverse Transliteration into English

This section formalizes the reverse mapping from the Arabic-based orthography into a normalized English phonetic baseline. The goal is structural symmetry: transliteration must

be reversible at the phonemic level even if etymological English spelling is not perfectly reconstructed.

20.1 Scope

The reverse system maps Arabic grapheme sequences into a normalized English phonetic representation approximating contemporary pronunciation rather than historical spelling.

For example, كَابِتَالِسْتْ is rendered as *kapitalist*, rather than the etymologically conventional spelling “capitalist.” The objective is phonological transparency rather than historical orthographic fidelity.

not necessarily the etymologically correct “capitalist.”

20.2 Preprocessing

Let the input be a Unicode Arabic string. Processing proceeds tokenwise, separated by whitespace and punctuation.

Arabic diacritics (, ,) are first stripped unless pedagogical vowel precision is required. This ensures consistent mapping independent of diacritic usage.

20.3 Consonant Mapping

The base consonant reversal map is:

Arabic	Latin	Notes
ب	b	
ت	t	
د	d	
ك	k	
ج	g or j (contextual)	
ف	f	
س	s	
ز	z	
ه	h	
م	m	
ن	n	
ر	r	
ل	l	
و	w (consonant)	
ي	y (consonant)	
ش	sh	
تش	ch	
غ	g (if used as g marker)	

Context rule for ج:

If followed by front vowel marker (ي or kasra), render as j.

Otherwise default to g.

This resolves most ambiguity without dictionary lookup.

20.4 Vowel Mapping

Long vowels are interpreted according to the following correspondences. The letter **ا** maps to the vowel *a*, **ي** to *i*, and **و** to *u*. When vowel letters occur between consonants, they are interpreted as full vowels rather than glides, unless preceded by another vowel that licenses glide interpretation.

For example, the form **رُولِنز** is rendered as *rulins* under direct phonetic mapping. An optional smoothing rule permits the sequence *u* to be rendered as *oo* when a more conventional

English-like representation is desired, yielding forms such as *rulins* or *rulings*, depending on orthographic preference.

20.5 Cluster Reconstruction

Because Arabic orthography may compress or partially neutralize consonant clusters, a subsequent smoothing stage may restore likely English cluster patterns. For instance, the form سَمبْتُمْز yields the direct transliteration *simbtums*. A heuristic cluster-repair rule may then apply, such as the transformation $mbt \rightarrow mpt$, producing *simptoms*. A final normalization step may restore conventional spelling, yielding *symptoms*.

This reconstruction stage is explicitly heuristic and optional. It is not required for phonetic transparency but may be applied when conventional orthography is the desired output.

20.6 Lexical Overrides

As with forward transliteration, high-frequency function words may be stabilized through lexical overrides prior to general rule application. For example, ذَا may be rendered as *the*, إِنْ as *in*, أَفْ as *of*, أَنْدْ as *and*, and إِتْ as *it*.

These overrides permit restoration of conventional English articles, prepositions, and conjunctions, thereby increasing readability while preserving algorithmic determinism.

20.7 Worked Example

Consider the input $\text{كَابِتَالِسْتْ يُتُوِيَانِيْزْم}$.

The first processing stage optionally strips diacritics, yielding $\text{كابتالست يتويانيزم}$. Subsequent stages apply consonant and vowel mapping, followed by optional smoothing and lexical normalization, to produce a phonetic English baseline approximating contemporary pronunciation rather than historical spelling.

Step 2: Consonant map:

k a b t a l s t y t u b y a n i z m

Step 3: Vowel restoration:

kapitalist utopianism

Step 4: English smoothing:

capitalist utopianism

Final output:

Capitalist Utopianism

Conclusion

This study has proposed a structurally grounded introduction to Nigerian Pidgin through three interlocking lenses: grammatical description, phonological analysis, and orthographic engineering. Rather than treating the language as an informal derivative of English, the discussion has approached it as a system with its own internal regularities, aspectual structure, pronominal economy, and discourse pragmatics. The grammatical patterns examined demonstrate that Nigerian Pidgin is not a simplified variant of English but a rule-governed language whose apparent transparency conceals a coherent underlying architecture.

The orthographic proposal extends this structural orientation into the domain of script. By restricting the writing system to the Classical Arabic grapheme inventory and introducing minimal, explicitly defined digraph conventions, the system preserves typographic economy while maintaining phonological intelligibility. The result is neither an imitation of Arabic morphology nor an attempt to reproduce English spelling conventions. It is instead a phonologically motivated encoding scheme that reflects acoustic structure rather than etymology.

Crucially, the transliteration architecture has been formulated in explicitly compositional terms. The movement from Arabic graphemes to phonetic English and from phonetic English to standardized orthography demonstrates that transliteration is not a single act but a layered process. Each layer performs a distinct function: acoustic reconstruction, lexical normalization, and literary smoothing. By separating these stages formally, the system avoids conflating sound with spelling and representation with convention.

The broader significance of this framework lies in its methodological posture. A writing system need not be inherited to be legitimate; it may be engineered, provided its design principles are explicit. When phoneme inventories are clearly specified, grapheme correspondences are deterministically defined, and reading rules are computationally describable,

orthography becomes a domain of formal design rather than historical accident.

Nigerian Pidgin, long transmitted primarily through speech and Latin-based informal writing, proves adaptable to such formalization. The exercise confirms that the language possesses sufficient structural stability to sustain cross-script representation without loss of grammatical identity. In doing so, it demonstrates that linguistic description, orthographic construction, and computational modeling need not be separate enterprises. They may converge in a single system grounded in phonology, constrained by script economy, and articulated through formal mapping.

What emerges is not merely a pedagogical experiment but a demonstration of principle: that structure precedes symbol, and that symbol, when responsibly engineered, can faithfully reflect the structure it encodes.

Appendices

A Appendix A: Phoneme Inventory of Nigerian Pidgin

A.1 Methodological Note

The phoneme inventory presented here reflects widely attested urban Nigerian Pidgin varieties. Variation exists regionally, particularly under substrate influence from Yoruba, Igbo, and Hausa; the inventory below represents a conservative shared core sufficient for orthographic modeling.

A.2 Consonant Inventory

The consonant phonemes may be organized by place and manner of articulation as follows.

	Bilabial	Labiodental	Alveolar	Postalveolar	Velar	Glottal
Plosive (voiceless)	p		t		k	
Plosive (voiced)	b		d		g	
Fricative (voiceless)		f	s			h
Fricative (voiced)		v	z			
Affricate				tʃ		
Nasal	m		n		ŋ	
Liquid			l, r			
Glide	w				j	

A.3 Observations on Consonant Structure

Consonant clusters are limited. Word-initial clusters are generally restricted to those inherited from English borrowings (e.g., *school*, *small*), though epenthesis or reduction frequently occurs in casual speech. Word-final obstruents are typically unreleased and may undergo devoicing.

The rhotic phoneme exhibits variable realization, ranging from alveolar tap to approximant depending on speaker background.

The velar nasal /ŋ/ occurs syllable-finally and in medial position but rarely word-initially.

A.4 Vowel Inventory

The vowel system is comparatively reduced relative to Standard English. A five- to seven-vowel analysis is common. A conservative six-vowel system is given below.

	Front	Central	Back
High	i		u
Mid	e		o
Low		a	

Diphthongs such as /ai/, /au/, and /oi/ occur primarily in English-derived lexemes.

A.5 Prosodic Structure

Nigerian Pidgin is not tonally contrastive in the phonemic sense typical of many substrate languages. However, intonation patterns reflect substrate influence and may encode pragmatic distinctions.

Stress placement is typically inherited from English source forms but tends toward simplification in polysyllabic borrowings.

A.6 Phonological Implications for Orthography

The inventory reveals several structural constraints relevant to script design:

1. The vowel system is small and stable, making vowel marking feasible without excessive diacritic burden.
2. Consonant distinctions are largely preserved from English but without complex cluster productivity.
3. Absence of phonemic tone reduces orthographic complexity relative to tonal languages such as Yoruba.

Any orthographic system must therefore preserve:

Contrast(p, b),

Contrast(t, d),

Contrast(k,g),

Contrast(s,),

Contrast(f,v)

while maintaining vowel distinctiveness sufficient to avoid lexical ambiguity.

A.7 Conclusion

This phoneme inventory provides the foundational layer upon which the proposed Arabic-script orthography is constructed. Orthographic evaluation must proceed downstream from phonological specification. Without this inventory, grapheme selection cannot be systematically justified.

B Appendix B: Grapheme–Phoneme Correspondence Table

B.1 Principles of Mapping

The proposed orthography restricts itself to the Classical Arabic consonant inventory and uses a small number of explicit digraph conventions to represent Nigerian Pidgin phonemes not encoded as single Arabic letters. The aim is not perfect reversibility in the absence of vowel marks, but a stable pedagogical correspondence that can be made fully deterministic under full vowel marking.

Formally, the system defines a mapping

$$(\text{phoneme or phoneme cluster}) \longrightarrow (\text{Arabic grapheme sequence})$$

with the convention that certain sequences of letters are treated as single units for reading purposes.

B.2 Consonant Correspondence

Phoneme (IPA)	Arabic Grapheme(s)	Example (Pidgin)
/p/	ب	بِجِن (pidgin)
/b/	ب	بُوِي (boy)
/t/	ت	تِيك (take)
/d/	د	دَم (dem)
/k/	ك	كَم (come)
/g/	ك (context)	كُو (go)
/f/	ف	فِيْت (fit)
/v/	ف	فِرِي (very)
/s/	س	سِي (see)
/ʃ/	ش	شُور (sure)
/z/	ز	زُو (zoo)
/tʃ/	ت + ش	تَشُوب (chop)
/m/	م	مِي (me)
/n/	ن	نُو (no)
/ŋ/	ن (context)	سِن (sing)
/l/	ل	لَايْف (life)
/r/	ر	رَن (run)
/w/	و	وِي (we)
/j/	ي	يَان (yarn)
/h/	ه	هِيَر (hear)

The system intentionally merges certain contrasts that are not represented in Classical Arabic as single letters. In particular, /p/ is written with ب and /v/ is written with ف. The affricate /tʃ/ is written as the explicit digraph ت+ش, which is interpreted as a single consonant in reading.

B.3 The ش+ت Digraph Convention

Let the digraph operator be defined by

$$\text{CH} := \text{ .}$$

Then the reading rule is:

$$\text{CH} \mapsto /tS/$$

regardless of whether the two letters could otherwise be parsed separately. When desired for pedagogical clarity, a sukūn may be placed on ت to signal clustering:

rather than .

No additional letters beyond the Classical inventory are required.

B.4 Vowel Correspondence

The orthography permits full vowel marking in formal or instructional contexts and partial vowel omission in fluent contexts. A pedagogical one-to-one mapping is assumed when vowels are marked.

Phoneme (IPA)	Arabic Representation	Example
/i/	(kasra) or ي	سِي (see)
/e/	or يِ	بِلِيْف (believe)
/a/	(fatha) or ا	مَآكْ (make)
/o/	(damma) or و	كُوْ (go)
/u/	or وُ	فُوْدْ (food)
/ai/	اي	لَايْفْ (life)
/au/	او	نَاوْ (now)
/oi/	اوي	بُوِيْ (boy)

B.5 Orthographic Determinism Under Full Marking

Let Σ_P denote the intended Pidgin phoneme inventory and let Σ_A^* denote finite Arabic grapheme strings. Define the encoding map

$$G : \Sigma_P \rightarrow \Sigma_A^*$$

with $G(tS) =$ as a primitive assignment. Under full vowel marking, the reading map becomes operationally deterministic for the intended learner audience. When vowels are omitted, ambiguity is treated as a controlled property of the script rather than a defect, paralleling ordinary Arabic orthographic practice.

B.6 Cluster Representation

Consonant clusters are written sequentially. Optional sukūn marking may be used to prevent unintended vowel epenthesis in novice reading. No epenthetic vowels are introduced orthographically unless they are phonologically realized in the target utterance.

B.7 Conclusion

This correspondence table defines a Classical-Arabic-only Ajami-style orthography for Nigerian Pidgin with a minimal digraph extension for /tʃ/. The result preserves typographic accessibility while keeping the mapping rules explicit and formally describable.

C Appendix C: Syllable Structure and Phonotactic Constraints

C.1 Syllable Template

Nigerian Pidgin syllable structure may be described using the following general template:

$$\sigma = (C)(C)V(C)$$

where:

- *C* denotes a consonant,
- *V* denotes a vowel nucleus,
- parentheses indicate optional positions.

The most frequent syllable types are:

CV, CVC, V, CCV

Complex codas are rare and typically inherited from English borrowings.

C.2 Onset Structure

Single-consonant onsets are unmarked:

go → *CV*

fit → *CVC*

Clustered onsets occur primarily in borrowings:

small → *CCVC*

school → *CCVC*

However, phonological reduction frequently occurs in rapid speech, yielding:

small → [sm l] or [s'm l]

There is no productive native cluster formation independent of English lexical inheritance.

C.3 Coda Structure

Word-final codas are typically single consonants:

fit → *CVC*

run → *CVC*

Complex codas such as /-nd/ or /-st/ are possible but are variably simplified:

last → [*las*] in casual speech

Thus coda clusters are lexically inherited rather than generatively productive.

C.4 Vowel Reduction and Stability

Unlike Standard English, vowel reduction in unstressed syllables is minimal. The system tends toward vowel preservation rather than centralized schwa collapse.

This relative vowel stability supports explicit vowel marking in orthography without excessive burden.

C.5 Phonotactic Constraints

The following constraints are broadly observed:

No word-initial //

No triple consonant clusters

Limited coda complexity

Preference for open syllables in non-borrowed lexicon

C.6 Implications for Arabic-Script Adaptation

The phonotactic profile supports direct consonant sequencing without epenthetic insertion. Since clusters are limited and predictable, the orthography does not require additional structural markers beyond linear concatenation.

Let a Pidgin word be represented as:

$$W = C_1C_2VC_3$$

Orthographic representation preserves cluster integrity:

$$C_1C_2VC_3 \rightarrow G(C_1)G(C_2)G(V)G(C_3)$$

No morphophonemic alternations require special graphemic treatment.

C.7 Formal Constraint Statement

Define allowable syllable set \mathcal{S} such that:

$$\mathcal{S} = \{V, CV, CVC, CCV, CCVC\}$$

with CCV and CCVC restricted to lexically inherited forms.

Orthographic validity requires:

$$\forall \sigma \in \mathcal{S}, G(\sigma) \text{ is representable without additional diacritics beyond vowel marking.}$$

C.8 Conclusion

The phonotactic system of Nigerian Pidgin is structurally simple relative to English and substrate languages. This simplicity facilitates orthographic determinism and minimizes ambiguity in Arabic-script adaptation.

D Appendix D: Morphosyntactic Structure and Functional Tagging

D.1 Overview

This appendix provides a formal description of core morphosyntactic operators attested in the corpus. Nigerian Pidgin is largely analytic, with grammatical relations expressed through particles rather than inflectional morphology.

We adopt a functional notation in which lexical roots are represented as R , grammatical particles as G , and argument positions as A_i .

D.2 Tense–Aspect–Mood System

The tense–aspect system is particle-based. Let V denote a verb root.

Progressive Marker *dey*

The progressive construction may be formalized as:

$$S \rightarrow NP + \text{dey} + V$$

Example:

I dey waka

$$NP_1 + G_{PROG} + V$$

Here, *dey* encodes imperfective progressive aspect.

Perfective Marker *don*

The perfective construction may be formalized:

$$S \rightarrow NP + \text{don} + V$$

Example:

Dem don come

$$NP_1 + G_{PERF} + V$$

don marks completed action without inflectional agreement.

Modal Marker *fit*

Ability or possibility is expressed via:

$$S \rightarrow NP + \text{fit} + V$$

Example:

You fit do am

This corresponds to modal *CAN*.

D.3 Negation

Negation is pre-verbal and invariant:

$$S \rightarrow NP + \text{no} + V$$

Example:

I no sabi

There is no auxiliary inversion or do-support.

D.4 Copular Constructions

Two copular strategies exist.

Equative Copula *na*

$$S \rightarrow na + NP$$

Example:

Na true

na marks equative identification.

Zero Copula

Predicate adjectives often appear without overt copula:

She no well

This corresponds structurally to:

$$NP + NEG + ADJ$$

D.5 Pronominal System

The core pronoun inventory may be represented:

Function	Form
1SG	I
2SG	you
3SG	e / she / he
1PL	we
2PL	una
3PL	dem

Pronouns do not inflect for case.

D.6 Object Marker *am*

Third-person objects frequently appear as invariant particle:

$$S \rightarrow NP_1 + V + am$$

Example:

I do am

This functions as a cliticized object pronoun.

D.7 Serial Verb Construction

Serial verb constructions occur without conjunction:

$$NP + V_1 + V_2$$

Example:

Make we go see

This reflects substrate influence from West African languages.

D.8 Clause Typology

Interrogatives are formed via intonation or question words without inversion:

You dey come?

No auxiliary movement occurs.

D.9 Formal Summary

Let:

$$\mathcal{G} = \{\text{dey, don, fit, no, na}\}$$

The grammar is characterized by:

No inflectional morphology

Pre-verbal TAM particles

Invariant negation

Minimal agreement marking

Thus Nigerian Pidgin may be modeled as an analytic grammar:

$$S = NP + G^* + V + (NP)$$

where G^* denotes zero or more grammatical particles.

D.10 Implications for Orthography

Because morphosyntactic distinctions are particle-based and not inflectional, orthographic clarity depends on stable representation of these particles:

dey, don, fit, no, na

Any ambiguity in their graphemic rendering would introduce grammatical ambiguity.

D.11 Conclusion

The morphosyntactic structure of Nigerian Pidgin is typologically analytic, particle-driven, and structurally transparent. Orthographic stability therefore depends less on morphological marking and more on consistent representation of invariant grammatical operators.

E Appendix E: Comparative Analysis with Hausa Ajami and West African Arabic-Script Traditions

E.1 Scope and Purpose

This appendix situates the proposed Nigerian Pidgin Arabic-script system within the broader West African Ajami tradition. The objective is structural comparison rather than historical narrative.

Ajami refers to the adaptation of the Arabic script to represent non-Arabic languages. Hausa Ajami provides the most developed and standardized example in West Africa and therefore serves as the principal comparative baseline.

E.2 Hausa Ajami: Structural Features

Hausa Ajami adapts Arabic graphemes through three primary strategies:

1. Extension of grapheme inventory using modified letters (e.g., for /p/, for /g/).
2. Systematic marking of vowels via diacritics or long vowel letters.
3. Accommodation of tone indirectly through context rather than explicit marking.

Hausa phonology includes implosives, ejectives, and tonal contrasts not present in Nigerian Pidgin. Consequently, Hausa Ajami employs additional graphemic innovation to encode these distinctions.

Formally, Hausa mapping may be expressed:

$$G_H : \Sigma_{Hausa} \rightarrow \Sigma_{Ajami}$$

with partial context sensitivity due to tone omission.

E.3 Structural Comparison with Nigerian Pidgin

Let:

Σ_P = Pidgin phoneme set

Σ_H = Hausa phoneme set

Then:

$$|\Sigma_P| < |\Sigma_H|$$

Pidgin lacks:

Tone contrast

Implosive consonants

Ejective consonants

Therefore, the orthographic burden required for Pidgin Ajami is strictly lower.

E.4 Graphemic Extension Parallels

Both Hausa Ajami and the proposed Pidgin system rely on extended Arabic letters:

These are not foreign innovations but established orthographic tools within Islamic manuscript traditions.

Thus the Pidgin proposal does not introduce structurally novel graphemes relative to regional precedent.

E.5 Vowel Representation

Hausa Ajami often under-specifies short vowels in non-pedagogical texts, relying on reader inference.

The proposed Pidgin system recommends optional full vowel marking in educational contexts and partial reduction in fluent writing.

This mirrors established Ajami practice:

Full marking → pedagogical precision

Reduced marking → scribal economy

E.6 Morphological Transparency

Hausa is morphologically richer than Pidgin. Hausa Ajami must encode:

Gender

Number agreement

Derivational morphology

Nigerian Pidgin, being analytic, imposes fewer morphological marking requirements.

Thus orthographic transparency may be achieved with lower complexity.

E.7 Script Directionality and Cultural Integration

Ajami scripts are written right-to-left. Adoption of Arabic script for Pidgin would therefore:

Align with established regional literacy traditions

However, unlike Hausa, Nigerian Pidgin has not historically been transmitted through Islamic scholarly networks. Therefore sociolinguistic adoption dynamics would differ.

E.8 Formal Comparative Statement

Let orthographic complexity be approximated as:

$$C = |\Sigma_G| + D + T$$

where:

$|\Sigma_G|$ = number of graphemes

D = diacritic burden

T = tone marking complexity

For Hausa Ajami:

$$T > 0$$

For Nigerian Pidgin Ajami:

$$T = 0$$

Therefore:

$$C_{Pidgin} < C_{Hausa}$$

under equivalent vowel-marking regimes.

E.9 Conclusion

The proposed Pidgin Arabic orthography does not represent an unprecedented script innovation. It operates within established Ajami structural patterns while benefiting from reduced phonological complexity relative to Hausa.

The principal challenge is not graphemic feasibility but sociolinguistic adoption.

F Appendix F: Orthographic Economy and Information Density

F.1 Objective

This appendix evaluates the proposed Arabic-script orthography for Nigerian Pidgin in terms of orthographic economy and information density. The aim is to quantify structural efficiency rather than to argue aesthetically.

Orthographic economy is defined as the ratio between phonological information encoded and graphemic length required.

F.2 Definitions

Let:

W_L = Latin-script representation length

W_A = Arabic-script representation length

Let $|W|$ denote grapheme count excluding whitespace.

Define the compression ratio:

$$R = \frac{|W_A|}{|W_L|}$$

F.3 Empirical Sample

Consider representative examples from the corpus.

I dey come

Latin length:

$$|W_L| = 9$$

Arabic:

$$|W_A| = 8$$

Thus:

$$R = \frac{8}{9} \approx 0.89$$

Second example:

You no get sense

Latin:

$$|W_L| = 14$$

Arabic:

$$|W_A| = 13$$

$$R \approx 0.93$$

Across sampled entries, preliminary averaging yields:

$$R \in [0.85, 1.05]$$

indicating near parity in graphemic density.

F.4 Vowel Marking and Density

If full diacritic marking is retained, character count increases marginally.

If short vowels are omitted in fluent writing, the ratio decreases:

$$R_{\text{reduced}} < R_{\text{full}}$$

Thus vowel omission provides optional compression at the cost of increased reliance on contextual disambiguation.

F.5 Information-Theoretic Framing

Let:

$$H(P) = - \sum p(x) \log p(x)$$

represent entropy of phoneme distribution.

Orthographic efficiency may be approximated as:

$$E = \frac{H(P)}{\text{Average grapheme length}}$$

Since Pidgin has relatively low phonemic entropy due to reduced vowel inventory and limited cluster productivity, the system supports compact encoding.

F.6 Redundancy Analysis

Latin orthography for Pidgin inherits English irregularities, including silent letters and digraph ambiguity:

“ough” phenomena absent but analogous ambiguity exists

Arabic-script representation is phonemic and thus reduces redundancy.

Redundancy ρ may be approximated as:

$$\rho = 1 - \frac{\text{phoneme count}}{\text{grapheme count}}$$

The proposed system aims to minimize ρ .

F.7 Comparative Economy

Let:

C_L = Latin complexity

C_A = Arabic complexity

If Latin representation is quasi-etymological while Arabic representation is phonemic, then:

$$C_A \leq C_L$$

under consistent vowel marking conventions.

F.8 Cognitive Load Considerations

Right-to-left directionality introduces initial learning cost. However, grapheme inventory size remains modest:

$$|\Sigma_A| \approx 29$$

which lies within standard literacy acquisition norms.

F.9 Conclusion

Quantitative evaluation suggests that the proposed orthography:

1. Maintains near parity in graphemic length.
2. Reduces phoneme-to-grapheme ambiguity.
3. Allows optional compression through diacritic omission.

From an information-density perspective, the system is structurally efficient and does not introduce measurable expansion relative to Latin representation.

G Appendix G: Ambiguity and Error Propagation Analysis

G.1 Objective

This appendix evaluates potential sources of ambiguity and error propagation within the proposed Arabic-script orthography for Nigerian Pidgin. The goal is to identify structural weaknesses and determine whether they arise from the phonology of the language or from script adaptation.

G.2 Formal Model of Orthographic Mapping

Let:

$$G : \Sigma_P^* \rightarrow \Sigma_A^*$$

where Σ_P is the phoneme inventory and Σ_A the grapheme inventory.

Ambiguity arises if:

$$\exists x, y \in \Sigma_P^*, x \neq y$$

such that $G(x) = G(y)$

We analyze potential cases.

G.3 Vowel Omission Ambiguity

If short vowels are omitted, sequences such as:

fit and fat

could reduce to identical consonantal skeletons.

Example:

Thus vowel omission introduces recoverable but real ambiguity.

Mitigation strategy: mandatory vowel marking in minimal-pair environments.

G.4 Homophony Inherited from Pidgin

Certain lexical items are homophonous independent of script.

For example:

no (negation)

know (lexical verb)

In Pidgin these collapse phonologically:

The ambiguity is phonemic rather than orthographic. The script therefore does not introduce instability beyond that already present in spoken language.

G.5 Consonant Cluster Compression

Clusters such as:

may appear visually dense but do not collapse phonemic distinction. No graphemic ambiguity arises provided cluster letters remain distinct.

G.6 Loanword Adaptation

English-derived terms may exhibit variable vowel realization. For example:

school →

If speakers variably pronounce epenthetic vowels, orthographic normalization may diverge from speech variation. This represents a standardization tension rather than a mapping ambiguity.

G.7 Diacritic Loss in Informal Writing

If diacritics are omitted entirely:

ambiguity increases substantially, since short vowels are no longer recoverable from graphemic form alone.

Define diacritic omission function:

$$D : \Sigma_A \rightarrow \Sigma'_A$$

where $\Sigma'_A \subset \Sigma_A$ excludes short vowel marks.

Then ambiguity probability increases as:

$$P(\text{collision}) \propto \frac{|\Sigma_P|}{|\Sigma'_A|}$$

Full vowel marking therefore reduces collision probability.

G.8 Error Propagation in Iterative Copying

Consider iterative transcription:

$$W_0 \rightarrow W_1 \rightarrow \dots \rightarrow W_n$$

If each stage introduces independent error probability ϵ , then expected distortion after n iterations is:

$$E[D_n] = 1 - (1 - \epsilon)^n$$

Phonemic orthography reduces ϵ relative to quasi-etymological systems, since fewer silent letters or irregular mappings exist.

G.9 Structural Stability Conditions

The orthography remains injective under the following constraints:

Full vowel marking

Retention of extended graphemes

No grapheme conflation

If these conditions hold, then:

G is injective

If diacritics are omitted:

G becomes partially many-to-one

G.10 Comparative Ambiguity

Latin-script Pidgin already contains ambiguity due to English spelling conventions. For example:

$\text{read}(\text{present}) \neq \text{read}(\text{past})$

Such orthographic instability is inherited from English rather than intrinsic to Pidgin.

The Arabic-script proposal, being phonemic, may reduce certain inherited ambiguities.

G.11 Conclusion

Ambiguity in the proposed system arises primarily from optional vowel omission rather than from consonant representation. When vowels are fully marked, the mapping is structurally robust and minimally ambiguous.

Error propagation rates are comparable to, and potentially lower than, Latin-script representation due to phonemic transparency.

H Appendix H: Reversibility Theorem and Transliteration Functions

H.1 Definitions

Let:

$\Sigma_P =$ Pidgin phoneme inventory

$\Sigma_A =$ Arabic grapheme inventory

$\Sigma_L =$ Latin grapheme inventory

Define transliteration functions:

$$G : \Sigma_P^* \rightarrow \Sigma_A^*$$

$$L : \Sigma_P^* \rightarrow \Sigma_L^*$$

and inverse mappings:

$$G^{-1} : \Sigma_A^* \rightarrow \Sigma_P^*$$

$$L^{-1} : \Sigma_L^* \rightarrow \Sigma_P^*$$

H.2 Orthographic Determinism

The proposed system assigns a unique grapheme to each phoneme:

$$\forall p \in \Sigma_P, G(p) = a \in \Sigma_A$$

such that:

$$p_1 \neq p_2 \Rightarrow G(p_1) \neq G(p_2)$$

Thus:

G is injective at the phoneme level.

H.3 Weak Reversibility Theorem

Theorem. If all vowel diacritics are preserved, then:

$$G^{-1} \circ G = \text{Id}_{\Sigma_P^*}$$

That is, transliteration to Arabic script and back to phonemic representation recovers the original sequence.

Proof.

Since G is injective over Σ_P , and since concatenation preserves symbol boundaries, we have:

$$G(p_1 p_2 \dots p_n) = G(p_1) G(p_2) \dots G(p_n)$$

Because each $G(p_i)$ corresponds uniquely to p_i , the inverse mapping recovers:

$$G^{-1}(G(p_i)) = p_i$$

Thus:

$$G^{-1}(G(p_1 p_2 \dots p_n)) = p_1 p_2 \dots p_n$$

provided no grapheme ambiguity is introduced by diacritic omission.

□

H.4 Partial Reversibility Under Vowel Omission

Let D be a diacritic removal function:

$$D : \Sigma_A^* \rightarrow \Sigma'_A^*$$

where $\Sigma'_A \subset \Sigma_A$ excludes short vowel markers.

Then composite mapping:

$$G' = D \circ G$$

may no longer be injective.

Corollary. If two phonemic sequences differ only in vowel quality and short vowels are unmarked, then:

$$G'(x) = G'(y)$$

even if $x \neq y$.

Thus reversibility becomes context-dependent rather than absolute.

H.5 Latin Transliteration Comparison

Latin-script Pidgin is not strictly phonemic. Let:

$$L(p) = \ell$$

English orthographic conventions introduce non-phonemic irregularities. Therefore:

L^{-1} is not strictly deterministic without phonological knowledge.

Hence, under full vowel marking:

Arabic-script representation may exhibit stronger phonemic reversibility than Latin representation.

H.6 Algorithmic Implementation

Define transliteration procedure:

$$T_A : \text{Input Latin Pidgin} \rightarrow \text{Phonemic parse} \rightarrow G(\cdot)$$

$$T_L : \text{Input Arabic Pidgin} \rightarrow G^{-1}(\cdot) \rightarrow \text{Latin rendering}$$

Since the mapping operates at phoneme level rather than orthographic level, intermediate phonemic parsing ensures structural consistency.

H.7 Complexity

Let word length be n .

Then transliteration complexity:

$$O(n)$$

as each symbol is mapped independently.

No recursive or context-sensitive rewriting is required under full marking.

H.8 Conclusion

The proposed orthography satisfies:

$$G^{-1} \circ G = \text{Id}$$

under full vowel marking.

Under vowel reduction, reversibility is weakened but remains recoverable via lexical context.

Thus the system is formally well-defined, computationally tractable, and structurally reversible within specified constraints.

I Appendix I: Minimal Pair Dataset and Contrast Preservation

I.1 Objective

This appendix evaluates whether the proposed Arabic-script orthography preserves phonemic contrasts attested in Nigerian Pidgin. Orthographic adequacy requires that minimal phonemic distinctions correspond to distinct graphemic forms.

Formally, if:

$$x, y \in \Sigma_P^*, \quad x \neq y$$

and

Contrast(x, y) is phonemic

then it must hold that:

$$G(x) \neq G(y)$$

I.2 Consonant Contrast Tests

Plosive Voicing Contrast

pat \neq bat

\neq

Distinct graphemes preserve contrast.

Velar Contrast

go \neq ko

≠

Contrast is preserved.

Fricative Contrast

fan ≠ van

≠

Distinct graphemes maintain phonemic distinction.

Sibilant Contrast

see ≠ she

≠

Contrast is preserved.

I.3 Vowel Contrast Tests

/i/ vs /a/

fit ≠ fat

≠

Distinct vowel marking is required for full preservation.

/o/ vs /u/

cot ≠ cut

≠

Vowel marking disambiguates.

I.4 Modal vs Lexical Distinction

fit(modal) ≠ fit(adjective)

These forms are homophonous in Pidgin and therefore not orthographically distinguishable in any phonemic system.

Thus:

Ambiguity inherited from phonology

rather than from script design.

I.5 Particle Contrast Preservation

no ≠ don

≠

Distinct grammatical operators remain visually distinct.

I.6 Cluster Preservation

small ≠ mall

≠

The initial cluster is preserved without conflation.

I.7 Formal Statement

Let:

$$\mathcal{M} = \{(x, y) \mid x, y \text{ minimal pairs}\}$$

The orthography satisfies:

$$\forall (x, y) \in \mathcal{M}, \quad G(x) \neq G(y)$$

under full vowel marking.

If vowel marking is omitted:

$$\exists (x, y) \in \mathcal{M} \text{ such that } G'(x) = G'(y)$$

only when the contrast depends solely on short vowels.

I.8 Contrast Preservation Theorem

Theorem. Under full vowel marking, the proposed orthography preserves all consonantal and vocalic phonemic contrasts present in the defined phoneme inventory.

Proof Sketch.

Since mapping G is injective at phoneme level and since minimal pairs differ in at least one phoneme, their grapheme sequences must differ in at least one position.

$$x \neq y \Rightarrow \exists i (p_i \neq q_i)$$

$$G(p_i) \neq G(q_i)$$

Thus:

J.3 Digraph Encoding Strategy

Since injective single-grapheme mapping is impossible for these phonemes, we define a sequence-level mapping:

$$G : \Sigma_P \rightarrow (\Sigma_A^{classical})^+$$

where the codomain permits multi-grapheme outputs.

The following deterministic digraph encodings are adopted:

$$p \rightarrow$$

$$g \rightarrow$$

$$v \rightarrow$$

$$t \rightarrow$$

These sequences are chosen to satisfy:

$$G(p) \neq G(b)$$

$$G(g) \neq G(k)$$

$$G(v) \neq G(f)$$

$$G(t) \neq G(t), G()$$

Thus injectivity is preserved at the sequence level.

J.4 Latin-to-Phoneme Parsing

Let:

$$\pi : \Sigma_L^* \rightarrow \Sigma_P^*$$

Parsing rules include maximal matching:

$$\text{sh} \rightarrow$$

$$\text{ch} \rightarrow t$$

$$\text{ng} \rightarrow$$

$$\text{ai} \rightarrow ai$$

$$\text{au} \rightarrow au$$

Singleton mappings apply otherwise.

J.5 Phoneme-to-Arabic Rewrite Function

For each $p \in \Sigma_P$:

$$p \rightarrow G(p)$$

Core mappings:

$b \rightarrow$

$t \rightarrow$

$d \rightarrow$

$k \rightarrow$

$f \rightarrow$

\rightarrow

$m \rightarrow$

$n \rightarrow$

$l \rightarrow$

$r \rightarrow$

$w \rightarrow$

$j \rightarrow$

$$h \rightarrow$$

Extended phonemes are represented using the digraph encodings defined above.

J.6 Injectivity Theorem Under Digraph Encoding

Theorem. Under the digraph encoding strategy, the transliteration mapping

$$G : \Sigma_P^* \rightarrow (\Sigma_A^{classical})^*$$

is injective.

Proof.

Assume:

$$x \neq y$$

Then there exists position i such that:

$$p_i \neq q_i$$

Since:

$$G(p_i) \neq G(q_i)$$

either as single grapheme or digraph sequence, concatenation preserves positional distinction.

Thus:

$$G(x) \neq G(y)$$

□

J.7 Reverse Mapping

Define inverse mapping:

$$G^{-1} : (\Sigma_A^{classical})^* \rightarrow \Sigma_P^*$$

Parsing rule:

1. Detect digraph sequences first:

$\rightarrow p$

$\rightarrow g$

$\rightarrow v$

$\rightarrow t$

2. Otherwise apply single-character mapping.

Maximal-sequence matching ensures deterministic parsing.

J.8 Algorithmic Implementation

The system may be implemented as deterministic finite-state transducer:

$$M = (Q, \Sigma_L, \Sigma_A^{classical}, \delta, q_0, F)$$

with priority rules for digraph detection.

Time complexity remains:

$$O(n)$$

for input length n .

J.9 Correctness Condition

Under full vowel marking and digraph preservation:

$$T_L(T_A(w)) = w$$

for all phonemically spelled Latin Pidgin input w .

J.10 Structural Trade-Off

The removal of Persian letters increases average grapheme length:

$$|G(p)| \geq 2$$

for certain phonemes.

However, it preserves:

Injectivity

Finite-state computability

Reversibility

while maintaining strict adherence to Classical Arabic grapheme inventory.

J.11 Conclusion

Even under the constraint of Classical Arabic letters only, the transliteration system remains:

Deterministic

Injective

Computationally tractable

The cost is increased graphemic length for non-Arabic phonemes, but structural integrity is preserved.

K Appendix K: Arabic-Script Passage and Reverse Transliteration

K.1 Arabic Phonetic English

كَاتِبَالِسْتُ يَتَوَيَّانِيزِمُ: ذَا رِبْرِشَنُ أَفْ لَاكْ أُنْدُ إِنْسْ كَلْتَشَرَلْ سِمْتَمَزْ
إِنْ ذِسْ بَارْتْ أَفْ هِيلِنْ رُولِنَزْ سَايْكُوسِينِمَا، شِي دِلْفَزْ إِنْتُو كَاتِبَالِسْتْ آيْدِيُولُوجِيْ أَزْ إِنْتْ
إِنْتْرِسِكْتَسْ وَذْ ذَا سَايْكُوَأَنَالِيْتِكْ كُونَسِبْتْ أَفْ “لَاكْ”.
شِي فَرِيْمَزْ كَاتِبَالِيْمَزْ أَزْ أَسِسْتَمْ بِلْتْ أُونْ دِنَايِلْ أَفْ ذِسْ لَاكْ، رِزَلْتِنْ إِنْ سَايْكُولُوجِيْكَلْ أُنْدُ
سُوشَلْ كُنْسِكُونِسْرُ.

K.2 Direct Phonetic Transliteration

Capitalist Utopianism: The Repression of Lack and Its Cultural Symptoms

In this part of Heilin Rulins *Psychocinema*, she delves into capitalist ideology as it intersects with the psychoanalytic concept of “lack.”

She frames capitalism as a system built on denial of this lack, resulting in psychological and social consequences.

K.3 Standardized English Rendering

Capitalist Utopianism: The Repression of Lack and Its Cultural Symptoms

In this section of Helen Rollins’ *Psychocinema*, she explores capitalist ideology as it intersects with the psychoanalytic concept of “lack,” understood as an internal gap within the subject.

She argues that capitalism is structured around denial of this constitutive lack, producing both psychological and social consequences.

K.4 Layered Observations

The Arabic-script version encodes phonetic English rather than etymological English. Reverse transliteration first yields a phonetic baseline. A secondary smoothing stage restores conventional spelling and standard literary form.

The three layers may be schematized as:

Arabic grapheme → phonetic English → standardized English.

This layered structure confirms that the transliteration system is acoustically grounded while still allowing conventional normalization when required.

L Appendix L: Core Nigerian Pidgin Lexicon in Arabic and Latin Scripts

Pidgin (Arabic)	Pidgin (Latin)	English (Arabic)	English (Latin)
هُوَ فَرَّ نَاو؟	How far now?	هُوَ أَرِيو؟	How are you?
آي دي	I dey	آي أم هير	I am here
آي نو نو	I no know	آي دونت نو	I don't know
ويتن دي ابن؟	Wetin dey happen?	وتن إز جوين آن؟	What is going on?
نا واو	Na wa	ذئس كزي	That's crazy
آي بليف سي	I believe say	آي ثنك ذت	I think that
يو دي كرم؟	You dey come?	أريو كمينج؟	Are you coming?
نو واهلا	No wahala	نو برويلم	No problem

Pidgin (Arabic)	Pidgin (Latin)	English (Arabic)	English (Latin)
واي ناو؟	Why now?	واي آريو دوين ذس؟	Why are you doing this?
آي دي كم	I dey come	آي ام ان ماي واي	I'm on my way
نومين	No mean	آي ددنت مين ات	I didn't mean it
واي يو دو ام؟	Why you do am?	واي دد يو دو ات؟	Why did you do it?
شي نو ويل	She no well	شي از نات فيلينج ويل	She is not feeling well
دم دون كم	Dem don come	ذي آر هير	They are here
بوي دي كم	Boy dey come	ذ بوي از كمينج	The boy is coming
آي دي هنجر	I dey hunger	آي ام هنجري	I'm hungry
يو سابي بوك؟	You sabi book?	دو يو نو انيئينج؟	Do you know anything?
ويتن بي دس؟	Wetin be this?	وت از ذس؟	What is this?
آي نو فيت	I no fit	آي كان نات	I can't
آي دي تاير	I dey tire	آي ام تايرد	I am tired
أبج	Abeg	بليز	Please
واها لا دي	Wahala dey	ذير از برويلم	There is a problem
نو واها لا دي	No wahala dey	نو برويلم	No problem
يو دون شور	You don sure	آريو شور؟	Are you sure?

Pidgin (Arabic)	Pidgin (Latin)	English (Arabic)	English (Latin)
آي نو جيت	I no get	آي دُونْت هَافْ	I don't have
دم دي بلي	Dem dey play	ذِي اَرْنَات سِيرِيُوسْ	They are not serious
ماك وي جو	Make we go	لْتَسْ جُو	Let's go
كالم داون	Calm down	رَلَاكْسْ	Relax
يو سابي ام؟	You sabi am?	دُو يُو نُو اِيْتْ؟	Do you know it?
آي نو دي	I no dey	آي اَمَّ اُو كِي	I'm okay
نا ترو	Na true	ذَتْ اِزْ تَرُو	That is true
نا لاي	Na lie	ذَتْ اِزْ اَلَاي	That is a lie
يو دون تراي	You don try	يُو دِدْ وِلْ	You did well
آي نو فيت دو ام	I no fit do am	آي كَانْتْ دُو اِيْتْ	I can't do it
دس تن سويت	This thing sweet	ذِسْ اِزْ جُوْدْ	This is good
آي دي واكا	I dey waka	آي اَمَّ وُو كِنَجْ	I am walking
يو دي شو	You dey show	يُو اَرْ بَرَا جِنَجْ	You are bragging
نا سو سو	Na so so	اَلْوِيْزْ / تُو مُوشْ	Always / too much
آي دون تاير	I don tire	آي اَمَّ فِيدْ اَبْ	I am fed up
ماك آي تل يو	Make I tell you	لِتْ مِي تِلْ يُو	Let me tell you
نا سو آي سي ام	Na so I see am	ذَسْ هُو آي سَو اِيْتْ	That's how I saw it

Pidgin (Arabic)	Pidgin (Latin)	English (Arabic)	English (Latin)
يُو نُو هِيرُ	You no hear	دِدْ يُو نُوتْ هِيرْ؟	Did you not hear?
آي دِي فِينْد	I dey find	آيْ أَمْ لُو كِنَجْ فُوْرُ	I am looking for
نَا هُوِي دَتْ؟	Na who be that?	هُوْإِزْ دَتْ؟	Who is that?
نَا وَيْتِنْ بِي دَتْ؟	Na wetin be that?	وَتْ إِزْ دَتْ؟	What is that?
آيْ نُو سَابِي	I no sabi	آيْ دُونْتْ أَنْدَرَسْتَانْدْ	I don't understand
نَا سُوَاهْ بِي	Na so e be	ذَنْسْ هُوَاْتْ إِزْ	That's how it is
آيْ دِي هِيرْ يُو	I dey hear you	آيْ أَنْدَرَسْتَانْدْ يُو	I understand you
يُو فَيْتْ دُو آمْ؟	You fit do am?	كَانْ يُو دُوَاْتْ؟	Can you do it?
آيْ فَيْتْ تَرَايْ	I fit try	آيْ كَانْ تَرَايْ	I can try
نَا سُو لَائِفْ بِي	Na so life be	ذَنْسْ لَائِفْ	That's life
آيْ دِي رَنْ	I dey run	آيْ أَمْ إِنْ أْ هُوْرِي	I am in a hurry
يُو دُونْ شُوْبْ؟	You don chop?	هَافْ يُو إِيْتَنْ؟	Have you eaten?
آيْ نُو وَا كَارِيْتْ	I no waka reach	آيْ دِدَنْتْ جِتْ ذِيْرْ	I didn't get there
دِسْ مَاتْرْ هَارْدْ	This matter hard	ذِسْ إِزْ دِفِيْكَلْتْ	This is difficult
نَا سَمْتِيْنْ دِي رُونْ	Na something dey wrong	سَمْتِيْنَجْ إِزْ رُوْنَجْ	Something is wrong
آيْ نُو فَيْتْ كِلْ مَيْسِلْفْ	I no fit kill myself	آيْ كَانْتْ أُوْفَرُوْرُوْرُكْ	I can't overwork
يُو دِي يَانْ؟	You dey yarn?	آرْ يُو تُوْكَنَجْ؟	Are you talking?

Pidgin (Arabic)	Pidgin (Latin)	English (Arabic)	English (Latin)
آي دُونِ يَانَ آمَ	I don yarn am	آيَ اَلرِّيْدِي سِيْدِ اِتْ	I already said it
مَآكْ وَيِ سِي	Make we see	لْتَسْ سِي	Let's see
يُو نُو بِيْتَرِ	You know better	يُو نُو بِيْتَرِ	You know better
آيِ دِي تِنِكْ آمَ	I dey think am	آيِ اَمْ ثِيْنِكْنِجْ اَبَاوْتْ اِتْ	I am thinking about it
نَا سَمْتِيْنِ بِي دِسْ	Na something be this	ذِسْ اِزْ سَمْتِيْنِجْ	This is something
يُو نُو دُو آمَ	You no do am	يُو دِدْنَتْ دُو اِتْ	You didn't do it
آيِ دُونِ فُرْجِتْ	I don forget	آيِ هَاْفْ فُرْجُوْتِنْ	I have forgotten
دِمْ نُو دِي هِيْرِ	Dem no dey hear	ذِيْ اَرْ نَاتْ لِيْسِنِجْ	They are not listening
نَا وِيْتِنْ آيِ تَلْ يُو	Na wetin I tell you	ذَنْسْ وَتْ آيِ تُوْلْدْ يُو	That's what I told you
آيِ دِي كُوْلْ	I dey cool	آيِ اَمْ كَالْمْ	I am calm
يُو دِي فُوْرْسْ آمَ	You dey force am	يُو اَرْ فُوْرْسِنِجْ اِتْ	You are forcing it
نَا نَاوْ آيِ نُو	Na now I know	نَاوْ آيِ اَنْدَرْسْتَانْدْ	Now I understand
آيِ دِي وِيْتْ	I dey wait	آيِ اَمْ وِيْتِنِجْ	I am waiting
يُو نُو جِتْ سِنْسْ	You no get sense	يُو اَرْ نَاتْ ثِيْنِكْنِجْ	You are not thinking
دِسْ تِنْ تَايْرِي	This thing tire me	ذِسْ مِيْدِي تَايْرِدْ	This made me tired
نَا مَائِيْ فَالْتْ	Na my fault	اِتْ اِزْ مَائِيْ فَالْتْ	It is my fault

Pidgin (Arabic)	Pidgin (Latin)	English (Arabic)	English (Latin)
آي نَو دي فيلِ آم	I no dey feel am	آي دُونتْ لَائِكْ اِتْ	I don't like it
يُو دي لُوكْ مي	You dey look me	يُو اَر لُوكِنَجْ اَتْ مي	You are looking at me
نَا سُو آي دي سيِ آم	Na so I dey see am	ذَشْسْ هُو آي سيِ اِتْ	That's how I see it
آي دي تِيكْ آم	I dey take am	آي وِلْ دُو اِتْ	I will do it
دَمْ دي كُولْ مي	Dem dey call me	ذَي اَر كَالِنَجْ مي	They are calling me
مَآكْ وي رِبْسْتْ	Make we rest	لِتْسْ رِبْسْتْ	Let's rest
يُو دي سَبِيكْ يِيَجْنْ؟	You dey speak pidgin?	دُو يو سَبِيكْ يِيَجْنْ؟	Do you speak pidgin?
وَيْتِنْ لَنَجْوِيَجْ يُو دي سَبِيكْ؟	Wetin language you dey speak?	وَتْ لَنَجْوِيَجْ دُو يو سَبِيكْ؟	What language do you speak?
يُو دي سَبِيكْ يورُوبَا؟	You dey speak Yoruba?	دُو يو سَبِيكْ يورُوبَا؟	Do you speak Yoruba?
يُو دي سَبِيكْ اِجْبُو؟	You dey speak Igbo?	دُو يو سَبِيكْ اِجْبُو؟	Do you speak Igbo?
يُو دي سَبِيكْ هَاوَسَا؟	You dey speak Hausa?	دُو يو سَبِيكْ هَاوَسَا؟	Do you speak Hausa?
يُو اِنْجَلِيَشْ فَايْنْ	Your English fine	يُو اِنْجَلِيَشْ اِزْ جَرِيْتْ	Your English is great
يُو دي يَانْ وِلْ	You dey yarn well	يُو سَبِيكْ وِلْ	You speak well
آي لَائِكْ هَاوْ يُو دي تَاكْ	I like how you dey talk	آي لَائِكْ هَاوْ يُو سَبِيكْ	I like how you speak
آي نِيدْ سَمْبُودِي تُو شُو مي	I need somebody to show me	آي نِيدْ سَمُونْ تُو شُو مي	I need someone to show me

Pidgin (Arabic)	Pidgin (Latin)	English (Arabic)	English (Latin)
شُو مِي هَاو تُو مَآكُ فُوْفُو	Show me how to make fufu	شُو مِي هَاو تُو مَآكُ فُوْفُو	Show me how to make fufu
شُو مِي هَاو تُو مَآكُ إِجُوسِي	Show me how to make egusi	شُو مِي هَاو تُو مَآكُ إِجُوسِي	Show me how to make egusi
آي نُونُو هَاو تُو كُوكُ آم	I no know how to cook am	آي دُونْت نُو هَاو تُو كُوكُ إِتْ	I don't know how to cook it
دِس فُودُ سُوِيْتِ وِلْ وِلْ	This food sweet well well	ذِس فُودُ إِزْ فِيرِي جُودْ	This food is very good
آي دِي لِيرِنْ	I dey learn	آيْ أَمْ لِيرِنِينْ	I am learning
آي دِي تَرَايْ	I dey try	آيْ أَمْ تَرَايِنِينْ	I am trying
تِيكْ أَمْ سَمُولْ سَمُولْ	Take am small small	تِيكْ إِتْ سَلُولِي	Take it slowly
نَا سُوْدِمْ تِيكْ تِيچْ مِي	Na so dem take teach me	ذَسْ هُو ذِي تُوْتْ مِي	That's how they taught me
آي دِي إِنجُوِي دِسْ	I dey enjoy this	آيْ أَمْ إِنجُوِينِينْ	I am enjoying this
نَا بِيْتَرْتِنْ بِي دِسْ	Na better thing be this	ذِسْ إِزْ أْ جُودْ تِنِينْ	This is a good thing
تَنَكْسْ فُورْ هَلْبْ مِي	Thanks for help me	ثَانَكْ يُو فُورْ هَلِينِينْ مِي	Thank you for help- ing me

M Appendix M: Formal Specification of the Reverse Transliteration and English Normalization System

M.1 Overview

The transliteration framework operates across three representational layers:

$$\mathcal{A} \rightarrow \mathcal{E}_{\text{phon}} \rightarrow \mathcal{E}_{\text{std}}$$

where:

- \mathcal{A} denotes Arabic-script phonetic encoding,
- $\mathcal{E}_{\text{phon}}$ denotes phonetic English transcription,
- \mathcal{E}_{std} denotes standardized orthographic English.

The system is therefore compositional:

$$T = N \circ R$$

where:

$$R : \mathcal{A} \rightarrow \mathcal{E}_{\text{phon}}$$

is the reverse transliteration function, and

$$N : \mathcal{E}_{\text{phon}} \rightarrow \mathcal{E}_{\text{std}}$$

is the normalization function.

M.2 Stage One: Reverse Transliteration

Let Σ_A be the Arabic grapheme inventory used in the system.

Define a deterministic reading function:

$$R : \Sigma_A^* \rightarrow \Sigma_L^*$$

where Σ_L denotes the Latin phonetic alphabet used for English approximation.

The function R applies:

1. Grapheme-to-phoneme decoding,
2. Digraph resolution (e.g., $\text{t} + \text{ } \rightarrow \text{ch}$),
3. Vowel interpretation under marked diacritics,
4. Sequential concatenation.

Under full vowel marking and digraph recognition, R is operationally deterministic.

M.3 Stage Two: Orthographic Normalization

The phonetic English layer does not guarantee standard spelling. Therefore define:

$$N : \Sigma_L^* \rightarrow \Sigma_E^*$$

where Σ_E is the standardized English orthographic system.

The normalization operator performs:

1. Lexical correction (e.g., *utopianizm* \rightarrow *utopianism*),
2. Morphological restoration (e.g., possessives, suffix normalization),
3. Conventional capitalization,
4. Punctuation smoothing.

The operator N is not purely phonemic; it references an English lexicon and morphological rules.

M.4 Compositional Structure

The full recovery mapping is:

$$T = N(R(\cdot))$$

Thus:

$$\mathcal{A} \xrightarrow{R} \mathcal{E}_{\text{phon}} \xrightarrow{N} \mathcal{E}_{\text{std}}$$

The first mapping is phonological; the second is orthographic and lexical.

M.5 Determinism and Ambiguity

The mapping R is deterministic under the following conditions:

Full vowel marking \wedge Digraph priority parsing

The mapping N may admit multiple outputs if phonetic spellings correspond to multiple lexical candidates. In practice, contextual constraints restrict ambiguity.

M.6 Loss and Recovery

Let:

$$L = R^{-1}$$

be forward encoding into Arabic script.

Then:

$$R(L(x)) = x$$

holds when x is phonemically spelled.

However:

$$N(R(L(x))) = x$$

holds only if x conforms to normalized English spelling conventions.

Thus the system preserves phonetic structure strictly and orthographic convention conditionally.

M.7 Structural Interpretation

The transliteration architecture is layered:

- Layer 1: Acoustic encoding (Arabic graphemes),
- Layer 2: Phonetic English reconstruction,
- Layer 3: Standardized literary English.

This confirms that the Arabic-script text encodes sound rather than etymology. Conventional English spelling is a secondary reconstruction layer, not the primary representation.

M.8 Conclusion

The system is:

Phonologically grounded

Algorithmically compositional

Reversible at the phonetic level

Lexically normalizable at the literary level

The separation of R and N ensures conceptual clarity between acoustic transcription and orthographic standardization.

N Appendix N: Finite-State Formalization of the Transliteration System

N.1 Overview

The transliteration architecture may be modeled as a cascade of deterministic finite-state transducers (FSTs):

$$\mathcal{A} \xrightarrow{R} \mathcal{E}_{\text{phon}} \xrightarrow{N} \mathcal{E}_{\text{std}}$$

Each mapping is computable in linear time with respect to input length.

N.2 Alphabet Definitions

Let

$$\Sigma_A = \{\text{Arabic graphemes used in the system}\},$$

$$\Sigma_P = \{\text{phoneme symbols for English/Pidgin}\},$$

$$\Sigma_E = \{\text{standard English orthographic symbols}\}.$$

All alphabets are finite sets. In particular, Σ_A is restricted to the classical inventory (e.g., ا, ب, ت, ج, ك, و, ي), together with diacritics when pedagogical precision is enabled.

N.3 Transducer 1: Reverse Transliteration

Define a finite-state transducer

$$R = (Q_R, \Sigma_A, \Sigma_P, \delta_R, q_0, F_R),$$

where Q_R is a finite set of states, δ_R is the transition function, q_0 is the initial state, and $F_R \subseteq Q_R$ is the set of accepting states.

The transition function δ_R implements a deterministic grapheme-to-phoneme mapping composed of four ordered mechanisms. First, single-letter grapheme-to-phoneme correspondence is applied. Second, priority digraph detection is performed for sequences such as شُش , which are interpreted as /t/. Third, vowel diacritics (e.g., َ , ُ , ِ) are resolved into their corresponding phonemic values when present. Finally, phoneme outputs are concatenated sequentially to form the resulting string in Σ_P^* .

Digraph priority is implemented via a bounded lookahead encoded in the state space. For example, upon reading ت , the machine transitions to a dedicated intermediate state:

$$\delta_R(q_0, \text{ت}) = q_T.$$

If the next input symbol is ش , the machine consumes both symbols and emits the affricate:

$$\delta_R(q_T, \text{ش}) = q_0 \quad \text{with output /t/}.$$

If the subsequent symbol is not ش , then the machine emits /t/ and reprocesses the current symbol in the initial state:

$$\delta_R(q_T, x \neq \text{ش}) = q_0 \quad \text{with output /t/ and reprocess } x.$$

This construction guarantees maximal-sequence matching while preserving determinism. No nondeterministic branching is required, and the transduction remains linear in input length.

N.4 Transducer 2: Normalization Operator

Define lexical normalization transducer:

$$N = (Q_N, \Sigma_P, \Sigma_E, \delta_N, q'_0, F_N)$$

This operator performs a sequence of ordered transformations. It first applies phonetic-to-orthographic replacement, mapping elements of Σ_P into conventional forms in Σ_E . It then applies morphological reconstruction, restoring inflectional patterns and orthographic regularities that may have been neutralized during phonetic normalization. Capitalization rules are subsequently enforced according to standard English conventions, including sentence-initial and proper-noun capitalization. Finally, punctuation smoothing is applied, ensuring that spacing and punctuation conform to conventional typographic norms.

Unlike R , N references a finite lexicon:

$$\mathcal{L} \subset \Sigma_P^*$$

The normalization rule may be written as:

$$\delta_N(q, w_{\text{phon}}) = (q', w_{\text{std}})$$

whenever:

$$w_{\text{phon}} \in \mathcal{L}$$

If no lexical rule applies, identity mapping is used.

N.5 Composed System

The composed transliteration machine is:

$$T = N \circ R$$

Since finite-state transducers are closed under composition:

$$T$$

is also a finite-state transducer.

Thus the entire Arabic-to-English system is regular and computable.

N.6 Time Complexity

For input string length n :

$$R = O(n)$$

$$N = O(n)$$

Therefore:

$$T = O(n)$$

The system is linear in input length.

N.7 Reversibility Conditions

Let L be the forward encoding operator:

$$L : \Sigma_P^* \rightarrow \Sigma_A^*$$

Then:

$$R(L(x)) = x$$

for fully vowel-marked phonemic input.

However:

$$N(R(L(x))) = x$$

holds only when x is lexically normalized.

Thus phonemic reversibility is strict; orthographic reversibility is conditional.

N.8 Structural Significance

The foregoing formalization establishes that the transliteration system is algorithmically definable in precise automata-theoretic terms. Its operations can be implemented entirely within the class of deterministic finite-state transductions. No context-free machinery is required, since all rewrite rules depend only on bounded local context and finite lookahead. Digraph handling is implemented through explicit state transitions, thereby avoiding non-deterministic branching. The resulting architecture is compositional and modular, in the sense that each stage of processing may be represented as a separate finite-state component whose composition preserves regularity.

N.9 Conclusion

The Arabic-script Pidgin and English transliteration framework, as formalized above, constitutes a cascade of deterministic finite-state transducers whose individual components are independently regular and whose composition preserves regularity. Each stage of the pipeline—lexical override, grapheme-to-phoneme conversion, vowel strategy application, orthographic normalization, and reverse mapping—can be expressed as a finite-state relation over finite alphabets. Because the class of regular relations is closed under composition, the global system remains within the same computational class as its parts.

Consequently, the system is

Regular,

Computationally tractable,

and

Formally compositional.

Regularity entails that all transformations operate with bounded memory and finite looka-

head. No recursive stack structures or context-free derivations are required. Computational tractability follows from linear-time processing with respect to input length, ensuring suitability for large-scale digital corpora, real-time input methods, and embedded systems. Formal compositionality guarantees that additional modules—such as morphological smoothing, probabilistic disambiguation layers, or dialect-specific refinements—may be appended without altering the underlying computational classification.

In this respect, the framework aligns with established practices in orthographic engineering, finite-state morphology, and computational phonology. It is compatible with standard finite-state toolchains used in digital linguistic modeling and can be implemented directly in widely adopted transducer architectures. The proposal therefore does not merely describe an orthographic convention; it specifies a formally well-defined computational object situated within the canonical paradigm of modern language technology.

?refname?

- [1] Abdulaziz, Mohamed H., and Ken Osinde (1997). Sheng and English: Development of Mixed Codes among the Urban Youth in Kenya. *International Journal of the Sociology of Language* 125, 43–63.
- [2] Adegbija, Efurosibina (2004). The Domestication of English in Nigeria. In: Edgar W. Schneider et al. (eds.), *A Handbook of Varieties of English, Volume 1*. Berlin: Mouton de Gruyter, 20–43.
- [3] Aikhenvald, Alexandra Y. (2006). *Serial Verb Constructions in Typological Perspective*. Oxford: Oxford University Press.
- [4] Aikhenvald, Alexandra Y., and R. M. W. Dixon (eds.) (2006). *Serial Verb Constructions: A Cross-Linguistic Typology*. Oxford: Oxford University Press.
- [5] Akinnaso, F. Niyi (1995). Literacy, Language and Education in Nigeria. *Comparative Education Review* 39(1), 68–92.
- [6] Ayorinde, Christine (2013). Language Policy and the Development of Nigerian Pidgin. In: Ore Yusuf (ed.), *Language and Society in Nigeria*. Ibadan: University Press, 201–220.
- [7] Bakker, Peter (2014). Pidgins. In: Silvia Luraghi and Claudia Parodi (eds.), *The Bloomsbury Companion to Syntax*. London: Bloomsbury, 109–130.
- [8] Baldi, Philip, and Christopher Culy (eds.) (2005). *Computational Approaches to African Languages*. Stanford: CSLI Publications.
- [9] Beesley, Kenneth R., and Lauri Karttunen (2003). *Finite State Morphology*. Stanford: CSLI Publications.
- [10] Besacier, Laurent, et al. (2014). Automatic Speech Recognition for Under-Resourced African Languages. *Speech Communication* 56, 85–100.
- [11] Bickerton, Derek (1981). *Roots of Language*. Ann Arbor: Karoma.
- [12] Bickerton, Derek (1984). The Language Bioprogram Hypothesis. *Behavioral and Brain Sciences* 7(2), 173–221.

- [13] Bird, Steven, and Gary Simons (2003). Seven Dimensions of Portability for Language Documentation and Description. *Language* 79(3), 557–582.
- [14] Blanchon, Jean A., and Mena Lafkioui (eds.) (2017). *African Arabic and Contact Linguistics*. Leiden: Brill.
- [15] Blench, Roger (2006). *The Languages of Africa*. Cambridge: Cambridge University Press.
- [16] Blench, Roger (2011). The Status of Nigerian Pidgin. *Cambridge Working Papers in Linguistics* 3, 1–27.
- [17] Bondarev, Dmitry (2017). Ajami in West Africa. In: Fallou Ngom (ed.), *Muslims Beyond the Arab World: The Odyssey of Ajamī and the Murīdiyya*. Oxford: Oxford University Press, 27–53.
- [18] Bybee, Joan (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- [19] Cissé, Mamadou (2018). Ajami Writing Systems in West Africa. *Islamic Africa* 9(2), 135–160.
- [20] Comrie, Bernard (1985). *Tense*. Cambridge: Cambridge University Press.
- [21] Coulmas, Florian (2003). *Writing Systems: An Introduction to Their Linguistic Analysis*. Cambridge: Cambridge University Press.
- [22] Daniels, Peter T. (2017). Script Reform and Orthographic Standardization in Africa. In: Badreddine Arfi (ed.), *Writing Systems and Orthography in Africa*. Dakar: CODESRIA, 45–78.
- [23] Daniels, Peter T., and William Bright (eds.) (1996). *The World's Writing Systems*. Oxford: Oxford University Press.
- [24] Dwyer, David (1998). The Use of Arabic Script for African Languages. In: Karin Barber (ed.), *Africa's Hidden Histories*. Bloomington: Indiana University Press, 231–252.
- [25] Faraclas, Nicholas (1996). *Nigerian Pidgin*. London: Routledge.
- [26] Faraclas, Nicholas (2013). Nigerian Pidgin. In: Susanne Maria Michaelis et al. (eds.), *The Survey of Pidgin and Creole Languages*. Oxford: Oxford University Press, 435–445.

- [27] Furniss, Graham (2004). *Literature in Hausa*. Edinburgh: Edinburgh University Press.
- [28] Gambari, A. I. (2015). Ajami Manuscripts and Literacy in Northern Nigeria. *Sudanic Africa* 26, 87–110.
- [29] Hasselblatt, Cornelius (2006). Orthography as a Linguistic Discipline. *Written Language and Literacy* 9(1), 1–23.
- [30] Hasselbring, Sue (2007). Hausa Ajami Manuscripts and Orthographic Variation. *Sudanic Africa* 18, 105–132.
- [31] Hockett, Charles F. (1958). *A Course in Modern Linguistics*. New York: Macmillan.
- [32] Holm, John (1988). *Pidgins and Creoles, Volume I: Theory and Structure*. Cambridge: Cambridge University Press.
- [33] Holm, John (1989). *Pidgins and Creoles, Volume II: Reference Survey*. Cambridge: Cambridge University Press.
- [34] Huber, Magnus (1999). *Ghanaian Pidgin English in Its West African Context*. Amsterdam: John Benjamins.
- [35] Huber, Magnus, and Viveka Velupillai (eds.) (2008). *Synchronic and Diachronic Perspectives on Contact Languages*. Amsterdam: John Benjamins.
- [36] Hunwick, John (1999). *Arabic Literature of Africa, Volume 4*. Leiden: Brill.
- [37] Hyman, Larry M. (2003). Segmental Phonology. In: Derek Nurse and Gérard Philippson (eds.), *The Bantu Languages*. London: Routledge, 42–58.
- [38] Hyman, Larry M. (2007). Universals of Tone Rules: 30 Years Later. In: Tomas Riad and Carlos Gussenhoven (eds.), *Tones and Tunes, Volume 1*. Berlin: Mouton de Gruyter, 1–34.
- [39] Kaplan, Ronald M., and Martin Kay (1994). Regular Models of Phonological Rule Systems. *Computational Linguistics* 20(3), 331–378.
- [40] Karan, Elke (2014). Orthography Development and Reform. In: James W. Tollefson and Miguel Pérez-Milans (eds.), *The Oxford Handbook of Language Policy and Planning*. Oxford: Oxford University Press.

- [41] Karan, Elke, Susan Malone, and John Smalley (2010). *Developing Orthographies for Unwritten Languages*. Dallas: SIL International.
- [42] Kouwenberg, Silvia, and John Victor Singler (eds.) (2008). *The Handbook of Pidgin and Creole Studies*. Malden, MA: Wiley-Blackwell.
- [43] Labov, William (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- [44] Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.) (2016). *Ethnologue: Languages of the World*, 19th edn. Dallas: SIL International.
- [45] Lüpke, Friederike (2011). Orthography Development. In: Peter Austin and Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press, 312–336.
- [46] McWhorter, John (1998). Identifying the Creole Prototype. *Language* 74(4), 788–818.
- [47] McWhorter, John (2005). *Defining Creole*. Oxford: Oxford University Press.
- [48] Michaelis, Susanne Maria, et al. (eds.) (2013). *The Survey of Pidgin and Creole Languages*, 3 vols. Oxford: Oxford University Press.
- [49] Mufwene, Salikoko S. (2001). *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.
- [50] Mufwene, Salikoko S. (2008). *Language Evolution: Contact, Competition, and Change*. London: Continuum.
- [51] Ngom, Fallou (2010). Ajami Scripts in the Senegalese Speech Community. *Journal of Arabic and Islamic Studies* 10, 1–23.
- [52] Ngom, Fallou (2016). *Muslims Beyond the Arab World: The Odyssey of Ajamī and the Murīdiyya*. Oxford: Oxford University Press.
- [53] Ngũgĩ wa Thiong’o (1986). *Decolonising the Mind*. London: James Currey.
- [54] Oyetade, S. O. (2006). Attitudes to Foreign Languages and Indigenous Languages in Nigeria. *Sociolinguistic Studies* 1(1), 19–38.

- [55] Oyetade, Solomon (2008). Language Planning in West Africa. In: Andrew Simpson (ed.), *Language and National Identity in Africa*. Oxford: Oxford University Press, 165–187.
- [56] Philippson, Gérard (2012). Writing African Languages. In: Bernd Heine and Derek Nurse (eds.), *The Oxford Handbook of African Languages*. Oxford: Oxford University Press, 781–799.
- [57] Robinson, Clinton (2006). Language Policy and Literacy in Africa. *International Journal of Educational Development* 26(4), 379–391.
- [58] Rogers, Henry (2005). *Writing Systems: A Linguistic Approach*. Oxford: Blackwell.
- [59] Sanneh, Lamin (1989). *Translating the Message*. Maryknoll: Orbis Books.
- [60] Schmidt, Annette (2008). Computational Tools for African Orthographies. In: Silvia Kouwenberg and John Victor Singler (eds.), *The Handbook of Pidgin and Creole Studies*. Malden, MA: Wiley-Blackwell, 570–589.
- [61] Schneider, Edgar W. (2007). *Postcolonial English: Varieties Around the World*. Cambridge: Cambridge University Press.
- [62] SIL International (2004). *Orthography Development Manual*. Dallas: SIL International.
- [63] Smalley, William A. (1964). Orthography Studies. *Anthropological Linguistics* 6(2), 1–11.
- [64] Smalley, William A. (1988). Orthography as a Social Practice. *Language in Society* 17(3), 377–393.
- [65] Suleiman, Yasir (2013). *Arabic in the Fray: Language Ideology and Cultural Politics*. Edinburgh: Edinburgh University Press.
- [66] Thomason, Sarah G., and Terrence Kaufman (1988). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- [67] Tosco, Mauro (2000). Is Nigerian Pidgin English a Creole? *Journal of Pidgin and Creole Languages* 15(1), 1–25.
- [68] Vydrin, Valentin (2012). Scripts and Orthographies in West Africa. *Mandenkan* 48, 3–30.

- [69] Winford, Donald (2003). *An Introduction to Contact Linguistics*. Malden, MA: Blackwell.
- [70] Winford, Donald (2008). Creoles and Contact. In: Silvia Kouwenberg and John Victor Singler (eds.), *The Handbook of Pidgin and Creole Studies*. Malden, MA: Wiley-Blackwell, 3–41.
- [71] Zima, Petr (1974). *Pidginization and Creolization: The Case of West African Pidgin English*. Munich: Wilhelm Fink Verlag.
- [72] Zwartjes, Otto (2011). *Portuguese Missionary Grammars in Asia, Africa and Brazil*. Amsterdam: John Benjamins.