

STANDARDISING NAIJÁ PHONOLOGY: A PREFERENCE GRAMMAR (PG) TECHNIQUE

¹Jacinta BENJAMIN-OHWODEDE & ²C.U.C. UGORJI

¹Department of English and Literature, University of Benin

¹Department of Linguistics Studies, University of Benin

DOI: <https://dx.doi.org/10.4314/jolls.v15i2.22>

Abstract

Nigerian Pidgin is a commonly used English-based creole, which has many social and regional varieties. These varieties make it difficult to develop a standard sound system, especially for spelling and literacy. This study examines the sound patterns of Nigerian Pidgin using Preference Grammar (PG). The aim is to show how PG can help identify and select the most appropriate sound forms from different varieties influenced by Nigerian indigenous languages. The study uses a qualitative method and draws samples from spoken data in the *Naija Corpus*. Selected speech patterns are analysed and presented in *tableaux* to show how preference procedures apply. The findings indicate that Nigerian Pidgin has many variations that make standardization challenging. This study submits that Preference Grammar (PG) provides a clear method of choosing consistent sound forms and developing a more stable phonological system for Nigerian Pidgin. These findings are useful for improving pronunciation and spelling systems and supporting computational applications of the language.

Keywords: Nigerian Pidgin, Phonology Standardisation Preference Grammar, Phonological Variation

Introduction

Nigerian Pidgin, also called Naijá, is a widely used English-based contact language in Nigeria and serves as a common language for people from different ethnic groups and is used in many social settings (Affia, 2023). Consequently, there are many variations of Naijá. Okere & Ajileye (2024) submit that in spite of being widely used, Naijá shows a lot of variation across regions and social groups and these variations are generally as a result of external factors. Significantly, variations of Naijá “demonstrate ethno-linguistic features of the areas in which they are spoken... Warri, Sapele, Calabar and Lagos” (Bob & Obiukwu, 2022, p.180). Hence, there are varieties of Naijá, such as, Lagos pidgin, Port-Harcourt Pidgin and Warri Pidgin. This is mostly observable in its phonological system and so it is often difficult to describe its sound clearly because of its influence by Nigerian indigenous languages. Consequently, there are challenges regarding developing a standard pronunciation and spelling systems. Earlier studies have examined its structure and variation but there is still no widely accepted method for selecting standard sound forms across varieties (see Mafeni, 1971; Elugbe &

Omamor, 1991 and Faraclas, 1996). More recent works examine Naijá as a lingua franca, its morphophonological aspects and syntactic features, (see Faraclas, 2012; Nlem & Epepe 2025; Attah et al., 2025; Akinbo & Ekiugbo, 2024; Ewata & Adebayo, 2024; Honkanen, M., 2022). Other works focus on computational pidgin studies (see Scholman et al., 2025; Lin et al., 2024). Thus, while the likes of Elugbe and Omamor laid the foundational sociolinguistic and structural framework for Naijá studies, recent studies built on this foundation and expanded into digital communication and computational modelling. Despite these advances, there still remains a gap regarding Naijá phonological representation and pronunciation norms, including a codified phonetic description across regional varieties. This lack of clear standard affects Naijá’s literacy development and further linguistic studies.

This study therefore addresses the aforementioned gap by applying the Preference Grammar approach as a formal method for selecting the most suitable phonological forms from different competing variants. It aims to examine phonological variation in Naijá and show how Preference

Grammar (henceforth, PG) can be used to select and organise sound patterns. The study also focuses on identifying differences across varieties, evaluating competing sounds and developing a structured phonological inventory. Using data from the *Naija Corpus*, this study demonstrates how this framework can help build a more consistent and organized sound system for Nigerian Pidgin. More significantly, it is also useful for computational linguistics, as it supports development of speech and language technology.

Empirical Studies

Naijá has been widely studied, most especially in terms of its social use and the need to develop it into a more standard language. This section reviews investigations carried out by several researchers on Naijá studies. Using different approaches and methodological resources different from this present study, some of them posit that because Naijá is widely used across Nigeria, there is a need to make it more stable and organised.

Agheyisi (1988) is of the opinion that developing its vocabulary and grammar will improve its status and usefulness. In supporting this suggestion, Amadi & Collin-Wegwu (2021) suggest that a standard spelling system be developed based on how easily they are understood, which is important for education. In the same vein, Mensah (2012) examines how the language is developing new grammatical forms. From a different perspective and in an attempt to demonstrate how Naijá is changing, Deuber et al. (2021) show that online communication helps create more regular spelling patterns. Even though variations still exist despite the regularities, (because people use both personal and shared spelling forms), Deuber et al. submit that standardisation may actually be happening naturally. This justifies the aim of this study, which is, suggesting a framework for the standardisation of Naijá. Because based on findings from Deuber et al. (2021), while some standards are emerging, there is still a lot of difference. Eke & Ochuba (2023) do not hesitate to point out challenges when it comes to Naijá literacy. Such challenges include negative attitudes towards Naijá and concerns about its harmful effect on learning English, especially in schools (Eke & Ochuba, 2023).

Although these studies provide information regarding the development of Naijá in terms of spelling and vocabulary, including social factors, only a few examine the sound system and how it affects standardisation. For instance, Ojarikre (2015) and Evbuonwan & Omozuwa (2025) confirm that Naijá has its own unique sound system, which is different from standard British English. However, they fail to account for the different phonological systems of Naijá's sub-varieties. These studies showcase how dynamic Naijá is, but they do not provide a method for handling variation in its sound system.

Okere & Ajileye (2024) supports the claim that variation in Naijá is influenced by indigenous Nigerian languages and different social contexts without proffering a clear method for selecting and standardizing sound pattern. Proposals like that of Balogun (2013) and Obi (2014) focus on spelling and policies regarding Naijá rather than the sound system itself. From the above reviews, it is evident that the phonological aspect of Naijá's standardization is still underdeveloped. This study addresses this gap by using PG as a method for selecting preferred sound patterns that will make up the Naijá phonological inventory. Unlike earlier studies, that only describe variation, this approach helps to evaluate and choose from a range of competing phonological forms.

Theoretical Framework

This study uses Preference Grammar (PG) (Ugorji, 2010; 2013) to analyse and standardise the sound system of Naijá. PG is a phonological model that helps to choose the most preferred sound forms within a language, in this instance, Naijá. It is designed to handle variation in real spoken language. PG is similar, but not exactly, like Prince & Smolensky's (1993) Optimality Theory (OT). This is because they use ranking to evaluate language forms but are different in certain ways. On the one hand, while OT uses abstract underlying forms and rules to produce outputs, PG focuses only on actual spoken forms, without using complex rules. Hence, the study submits that PG is more practical as it is based on real language use. A significant feature of PG is that it uses parameters. these are similar to rules or guidelines used to judge different sound forms. The parameters are ranked,

meaning that some are more important than some, as we will see in this study's analysis. The parameters, which can be likened to OT's constraints, are flexible; they can be broken when necessary. This allows PG to compare different phonological forms and choose the most appropriate ones based on real usage. Unlike traditional approaches, PG does not compare Naijá to Standard British English. Rather, it evaluates the language on its own terms and selects preferred forms from within the language itself. Thus, this framework makes it applicable for the standardisation of Naijá phonological inventory of speech sounds, which has many regional varieties.

PG uses a set of parameters (Ugorji, 2013) to analyse sound segments and rhythm in spoken data. As mentioned earlier, these parameters work together in a ranking system, where more important ones guide the final choice of forms. Because the parameters are flexible, PG can handle variations while still identifying stable and more preferred patterns. In this study, PG is applied to phonological data from Naijá. It is used to compare different sound patterns and select those that can serve as standard or normative forms. The six parameters include:

International Acceptance (**I-Ac**)

Contrastiveness (**Contr**)

Frequency (**FreQ**)

Disambiguity (**DisaM**)

Phonetic Simplicity (**Pho-S**)

Pedagogical Convenience (**Ped-C**)

Preference Grammar (PG) Operational Terms

Several operational phrases and symbols are often used to assess and choose preferred phonological patterns in the Preference Grammar (PG) process. The following definitions apply to these terms:

a. *Attested Forms (Input Units)*: These are the phonological forms seen in speech that occurs spontaneously. In the PG framework, they stand in for the unprocessed data or candidates for assessment.

b. *Units of Output (Preferred Units)*: The units that the PG mechanism deems ideal are the preferred units. They represent the normative forms of the language and reflect the formal or standard components of the phonological inventory.

c. *Superfluous Components*: These variants

exhibit dialectal or sociolinguistic variance. They are important for descriptive and comparative reasons even if they are not included in the standard inventory.

d. *Mark of Violation (*)*: An applicant who has been rejected during the evaluation process due to a violation of one or more criteria is indicated by the asterisk symbol.

e. *Indicator of Preferred Candidate (→)*: The candidate form that is preferred or chosen during the evaluation is indicated by the bold arrow sign. The normative phonological inventory includes this form.

Methodology

This study uses a qualitative research design. Thus, the analysis describes and analyses real speech data, rather than the use of numbers or experiments. Data were taken from Naija Corpus an online collection of spoken Naijá, specifically, *NaijaSynCor*, a spoken corpus of Nigerian Pidgin consisting of approximately 400,000-500,000 words. This corpus is based on recorded speech of 320 speakers from 11 different survey locations in Nigeria and includes natural conversations such as personal stories discussions and everyday interactions. This study makes use of data from this corpus because it covers a wide range of speakers from different ethnic groups across the six-geopolitical zones, including Igbo, Hausa, Urhobo, Yoruba, and Ibibio amongst others. Also, it contains different topics and speaking situations, reflecting how Naijá is actually used in everyday situations.

This makes the corpus appropriate for studying variations in Naijá. In this study, data was drawn from 321 audio texts/conversations involving all 320 speakers. Although 43 words form the dataset for this study's analysis, 7 illustrative examples are included in the tableaux to demonstrate the identified patterns. Since the study is largely demonstrative, a limited dataset was considered adequate for the study's analysis and they capture a significant proportion of the phonemic inventory of Naijá. Regarding data collection, speech sounds are elicited from selected verbal items that show variations that represent differences across speakers and regions. During the study's analytical procedure, each selected item is isolated and grouped based on similar sound patterns. This makes it easier to compare different phonological forms.

The analysis is done using PG. the selected sound patterns are presented in *tableaux*, which is the format adopted the PG framework. Each *tableau* contains different phonological forms (variants); the parameters used for evaluation and the use of an asterisk indicating the preferred form based on ranking. Thereafter, the outcome of each analysis as captured in the *tableaux* is described.

Data Presentation and Analysis

This section shows how PG is used to select sound forms from Naija Corpus . The study uses transcription symbols adapted from David Jowitt (1991), which reflect the sound system of English and some Nigerian indigenous languages. The sound patterns taken from Naija Corpus are treated as possible input forms. The different forms are placed in *tableaux*, where they compete based on the six parameters of PG. as mentioned earlier, these parameters are ranked, so some are more important than others. Forms that meet the more important parameters are selected as the preferred candidates, while others are eliminated. In this way, this study’s analysis helps to identify and build a consistent set of sound patterns for Naijá.

Tableau 1: <palaver>

palava	*No I-Ac	*No ContR	*No FreQ	DisaM	*Pho-S	*Ped-C
/palava/			*			
→/kpalava/					*	

Tableau 1 above has two phonemes as inputs, the doubly articulated voiceless labio-velar stop and the voiceless bilabial stop /kp, p/ competing for preference, having the distinctive feature definable as consonantals. They are spelt within the context of /p/ in Naijá lexical items such as: <kpai>, <kamkpe>, <kpotoki>, <kpangolo>, <kpekere> and <kpoh-kpoh-kpoh> meaning *to die/kill, very well, Portuguese, metal can, fried plantain chips* and *a knocking sound*, respectively. The winning or preferred candidate in this instance, is /kp/, which automatically becomes a member of the standard or formal inventory because it violates a low-ranking parameter. The consonant input, /p/ is categorized as an allophone (still within the formal inventory) of the winning candidate, /kp/ because it violates the highest ranked parameter namely, *No FreQ compared to the input, /kp/ which violates a lower ranked parameter /*Pho-S/.

Tableau 2: <quench>

kwench	*No I-Ac	*No ContR	*No FreQ	DisaM	*Pho-S	*Ped-C
/kwenʃ/	*					
→ /kwenʃ/					*	

Tableau 2 demonstrates that there is the tendency for Naijá speakers to alternate /ʃ/ and /s/; voiceless affricate and the voiceless palate-alveolar fricative. This is seen in some Naijá words like: <choke>, <shege>, <chop> <winch>, <chook> and <shikena> meaning *surprising, difficulty, eat, witch, pierce, and that’s all* respectively. Having the distinctive feature definable as +obstruent, /s vs ʃ/, they are spelt within the context of <sh- or ch-> in the above Naijá lexical items. The first form /ʃ/ has violated high ranking constraint *No I-

Ac and belongs to the non-standard regional minority forms or the outer standard. The consonant /tʃ/ is thus, the winning or preferred candidate.

Tableau 3: <now!>- an exclamation for emphasis

Nah!	*No I-Ac	*No ContR	*No FreQ	*DisaM	*Pho-S	*Ped-C
→ /næ/				*		
/naʊ/				*	*	*
/nɔ/	*			*		

Tableau 3 above has three inputs, /a aʊ o/ competing for preference, having the distinctive feature definable as sonorants. They are spelt within the context of <ou> in Naijá lexical items such as: <domot>, <how fah>, <ah̄t/out>, <sidon> and <witat> door-mouth, how is everything, out, sit down and <without> respectively. The winning or preferred candidate is /a/, which automatically becomes a member of the standard/formal inventory because it violates just one low-ranked parameter. The Nigerian English monophthong, /o/ is categorized as an allophone (still within the formal inventory) of the winning candidate, /a/ because it violates just two low-ranked parameters namely, *No I-Ac and *DisaM. The input, /aʊ/ violates the highest-ranked parameter, *No I-Ac, giving /o/ an advantage over it. It therefore falls within the informal standard or the outer circle.

Tableau 4: <nothing>

Notin	*No I-Ac	*No ContR	*No FreQ	*DisaM	*Pho-S	*Ped-C
→ /nɔ̄tɪn/						
/nɔ̄ɐ̄ɪn/			*			
/nɔ̄rɪn/	*					

Tableau 4 has three consonant forms as inputs, the voiceless alveolar stop, the voiced interdental fricative and the voiced approximant roll /t ɐ r/ competing for preference, having the distinctive features of the first two, definable as obstruents and the third as a sonorant. They are spelt within the context of <th-> in the Naijá lexical items namely: <son̄tin>, <notin>, <tink>, <wit> and <tanks> meaning: something, nothing, think, with and thanks. The winning or the preferred candidate is the voiceless alveolar stop /t/, which inevitably becomes a member of the standard inventory because it does not violate any parameter. The voiceless interdental fricative /θ/ is categorized as an allophone of the winning candidate, /t/ in the context of the above words because it does not violate the two highest ranked parameters.

Tableau 5: <sit down>

Sitdon	*No I-Ac	*No	*No FreQ	*DisaM	*Pho-S	*Ped-C
--------	----------	-----	----------	--------	--------	--------

		ContR				
/sitdɔn/	*					
/sitdaun/			*			
→ /sitdan/						

Tableau 5 has three vowel forms as inputs, an open-back vowel, a closing glide and an open-front vowel /ɔ ə/ competing for preference, all having the distinctive features of sonorant. They are spelt within the context of <a> in the Naijá lexical items namely: <dan>, <cant>, <rand>, <pand> and <grand> meaning: down, count, round, pound and ground. The winning candidate is the front vowel /æ/, which inevitably becomes a member of the standard inventory because it does not violate any parameter. The back vowel /ɔ/ is categorized as an allophone of the winning candidate, /æ/ in the context of the above words because it does not violate highly ranked parameters.

Tableau 6: <them>

Dem	*No I-Ac	*No ContR	*No FreQ	*DisaM	*Pho-S	*Ped-C
→/dɛm/						
/ðɛm/			*			
/ʒɛm/	*					

Tableau 6 has three consonant forms as inputs, the voiced alveolar stop, the voiced inter-dental fricative and the voiced palate-alveolar fricative /d ð ʒ/ competing for preference, all having the distinctive features of obstruent. They are spelt within the context of <a> in the Naijá lexical items namely: <dey>, <dis>, <dat>, <togeda> and <den> meaning: they, this, that, together and then. The winning or the preferred candidate is the voiced alveolar plosive /d/, which certainly becomes a member of the standard inventory for the reason that it does not violate any parameter. The voiced inter-dental fricative /ð/ is considered as an allophone of the winning candidate, /d/ in the context of the above words because it does not violate the two highest ranked parameters.

Tableau 7: <crazy>

Craze	*No I-Ac	*No ContR	*No FreQ	*DisaM	*Pho-S	*Ped-C
→ /kres/						
/kreis/			*			

Tableau 7 has two vowel forms as inputs, an open-back vowel and a glide /e ei/ competing for preference, having the distinctive feature, sonorants. They are often indicated within the context of <e, ey, ay> in the

Naijá lexical items namely: <dey>, <kampe>, <tay-tay> and <werey> meaning: fine/okay, well/strong, long time ago and mad person. The winning or the preferred candidate is the front vowel /e/, which automatically becomes part of the standard inventory because it does not violate any parameter. The closing diphthong /ei/ becomes as an allophone of the winning candidate, /e/ in the context of the above words because it does not violate any highly ranked parameters. As established in *Tableaux* 1-7, the evaluation procedures of selected phonological units, does not necessarily “need to be exhausted, if higher ranked parameters are already violated by a given candidate in the competition and/or a winning candidate has clearly emerged” (Ugorji, 2010, p. 80). The application of PG helps in selecting candidates that are preferred for inclusion into the Naijá phonological inventory. Below, is a catalogue of preferred candidates of vowels for the Naijá phonological inventory.

Table 8: Naijá Vowels

Standard Monophthongs/Allophones
i:/ɪ
æ
e
ɛ
o
ɔ

Other monophthongs include, /a/ and /ɔ/. The Naijá diphthongs include /aɔ/, ɔɪ/ and /aɪ

Table 9: Naijá Consonants

Standard Consonants/Allophones
f
s/ʃ
v
t/d
ɖʒ/tʃ

Note that other consonants that form part of the Naijá phoneme inventory include: /b p g k ŋ m n k̄p ḡb w l r h ɲ z ʒ j/.

Conclusion

This study examines phonological variation in Nigerian Pidgin, Naijá and how Ugorji's (2013) Preference Grammar (PG) is used to select sound patterns. The study confirms that Naijá has phonological variation across its different varieties as observed by earlier studies. However, while previous studies focused more on spelling and vocabulary, this study focuses on the sound system of Naijá. *Tableaux* 1-7 show that once a phonological form has already failed important parameters, a winner emerges. PG allows different

forms to be compared and helps in selecting those that best meet higher-ranked parameters. through this technique, this study identifies a set of preferred vowels and consonants for Naijá. These findings suggest that even with variations within the language, stable and preferred sound patterns can be organised into a phonological inventory. This study provides a more stable and structured phonological system and also supports standardization in terms of spelling development, teaching and language technology.

References

- Affia, P. (2025). Nigerian pidgin English: the identity of a Nigerian away from home. *Working Papers in Applied Linguistics & Linguistics at York*, 3. 1-16. <https://doi.10.2571/2564-2855.23>
- Agheyisi, R. (1988). *The standardization of Nigerian Pidgin English*. *English World-Wide*, 9(2). Pp. 227-241. <https://doi.org/10.1075/eww.9.2.06agh>
- Akinbo, S. K. & Ekiugbo, P. O. (2024). Iconicity as the motivation for morphophonological metathesis and truncation in Nigerian Pidgin. *De Gruyter*. <https://doi.org/10.1515/opli-2024-0013>
- Amadi, L. C., & Collins-Wegwu, V. F. (2021). Nigerian Pidgin: an effective communicative device for vocabulary development. *Icheke Journal of the Humanities*, 19(1). 101-115.
- Attah, M. P., Salisu, I. & Yecho, E.N. (2025). IS Nigerian pidgin a language? A sociolinguistic enquiry. *Jalingo Journal of Language and Literary Studies (JAJOLLS)*, 9 (3). 45-56.
- Balogun, T. A. (2013). In defence of Nigerian Pidgin. *Journal of Languages and Culture*, 4(5). 90-98. <https://doi:10.5897/JLC12.007>
- Bob, O. P. & Obiukwu, E. N. (2022). Exploring the linguistic status of the Nigerian pidgin. *Preorcjah*, 7. 173-186. <https://www.ezenwaohaetorc.org>
- Deuber, D., Shakir, M. & Oyebola, F. (2021). Orthographic standardization of Nigerian pidgin in computer-mediated communication revisited.
- Eke, R. A. & Ochuba, O. O. (2023). The status, prospects and challenges of using Nigerian Pidgin as medium of instruction in Nigerian schools for equalisation of educational opportunities. *Journal of Education in Developing Areas*, 13(2). 322-332. <https://journal.journalsplace.org/index.php/JEDA>
- Elugbe, B. O. & Omamor, A. P. (1991). *Nigerian Pidgin: Background and prospects*. Heinemann Educational Books
- Evbuonwan, O. O. & Omozuwa, E. V. (2025). Towards a phonetic-based alphabet for Nigerian Pidgin. *Awka Journal of Linguistics and Languages*, 14 (2021), 1-20.
- Ewata, T.O. & Adebayo, M. O. (2024). Trends in Nigerian Pidgin orthography on digital platforms. *Journal of the Nigerian Lagos Project*, 6 (2). <https://doi.org/10.60787/jnlp.v6i2.70>
- Faraclas, N. (1996). *Nigerian Pidgin*. Routledge.
- Faraclas, N. (2012). The pronoun system in Nigerian pidgin: a preliminary study. In J. Cheshire, (Ed.). *English around the world: Sociolinguistic perspectives*. Pp. 509-518.

- <https://doi.org/10.1017/CB09780511611>
- Honkanen, M. (2022). “This word no get concrete meaning oo”: pragmatic markers in Nigerian online communication, *Journal of Pragmatics*, 202. Pp.93-104.
<https://doi.org/10.1016/j.pragma.2022.10.015>
- Jowitt, D. (1991). *Nigerian English Usage*. Longman.
- Lin, P., Scholman, M., Saeed, M. & Demberg, V. (2024). Modeling orthographic variation improves NLP performance for Nigerian Pidgin. <https://doi.org/10.48550/arXiv.2404.18264>
- Mafeni, B. (1971). Nigerian Pidgin. In J. Spencer, (Ed.). *The English Language in West Africa*. Longman. Pp. 95-112.
- Mensah, E. (2012). Grammaticalization in Nigerian pidgin. *Ikala, Revista De Lenguaje Y Cultura*. 17(2). 167-179. <https://doi.org/10.17533/udea.ikala.10748>
- Nlem, A.O. & Epepe, U. D. (2025). Nigeria pidgin as lingua franca: emerging communication trend in the Anthropocene. *Zenodo*, 1. <https://doi.105281/zenodo.17810612>
- Obi, E. I. (2014). Language attitude and Nigerian Pidgin. *IJAH*, 3(4). 34-46.
<https://dx.doi.org/10.4314/ijah.v3i4.3>
- Ojarikre, A. (2015). A contrastive study of the segmental phonology of Nigerian Pidgin English and standard English. *Journal of Literature, Language and Linguistics*, 12, 8-16.
- Okere, E.C. & Ajileye, M.O. (2024). Variety differentiation in Nigerian pidgin: a sociolinguistics study. *The international Journal of Humanities and Social Studies*, 12(4). 65-69.
<https://doi:10.24940/theijhss/2024/v12/i4/HS2404-012>
- Prince, A., & Smolensky, P. (1993). *Optimality theory: constraint interaction in generative grammar*. Rutgers University Centre for Cognitive Sciences.
- Schoolman, M.J., Marchal, M., Brown, A. & Demberg, V. (2025). DiscoNaija: a discourse-annotated parallel Nigerian Pidgin-English corpus. *Language Resources & Evaluation*, 59. Pp. 3597-3633.
<https://doi.org/10.1007/s10579-025-09850-3>
- Ugorji, C.U.C. (2010). *Nigerian English Phonology*. Peter Lang.
- Ugorji, C.U.C. (2013). New norms in English teaching and learning: the preference grammar approach, *Journal of the English Scholars Association*, 24 (2). 197-210.