Laura McPherson

A Grammar of Tommo So

Mouton Grammar Library

Edited by
Georg Bossong
Bernard Comrie
Matthew Dryer
Patience L. Epps

Volume 62

Laura McPherson

A Grammar of Tommo So

DE GRUYTER
MOUTON

ISBN 978-3-11-030092-5
e-ISBN 978-3-11-030107-6
ISSN 0933-7636

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Typesetting: RoyalStandard, Hong Kong
Printing and binding: Hubert & Co. GmbH & Co. KG, Göttingen
♾ Printed on acid-free paper
Printed in Germany

www.degruyter.com

For Ramata Ouologuem

Acknowledgments

It has been a journey preparing this grammar, a journey I could not have possibly undertaken on my own. I owe the existence of this volume to many people – consultants, friends, colleagues, editors – as well as to those reviewers who read grant applications and thought that this work was worthwhile.

First and foremost, I want to thank my consultants, Ramata Ouologuem, Sana 'M. le Maire' Ouologuem, Issa Toloba, Zakaria Ouologuem, Ajumare Guindo, Endekiye Ouloguem, and Ende Ouologuem for their patience and good humor in teaching me their language. I would also like to extend my thanks to the people of Tongo-Tongo for their hospitality. I am grateful to my non-Tommo friends in Mali as well, particularly Minkailou Djiguiba, Seydou Moro, Oumar Koné, and Salif Morogoye, for helping me so much in my day-to-day life in Mali. Thank you also to the linguists at the Institut National des Langues – Abdoulaye Barry in Bamako for fruitful discussions on the Dogon languages.

This work would not have been possible without Jeffrey Heath, who has trained and supported me throughout this project. Seeing firsthand his hardwork and dedication in the field has shaped me as a fieldworker; if I can accomplish even a small fraction of what he has, I will consider my career a success. Thank you also to my colleagues in the Dogon Languages Project, Kirill Prokhorov, Steve Moran, Brian Cansler, Vadim Dyachkov, and especially Abbie Hantgan, who have kept me company in the field. You are my fieldwork family.

This manuscript has benefitted from the comments of many colleagues, including Mary Paster, Laura Kalin, Kevin Ryan, Byron Ahn, Natasha Abner, and Denis Paperno, as well as my tireless advisors Bruce Hayes and Russell Schuh, who have taken time out of their already busy schedules to read portions of this grammar and suggest improvements. Any errors that remain are my own.

I am also very grateful to my contact editor, Matthew Dryer, for his careful reading of the manuscript. This final product has greatly benefitted from his wisdom on matters of typology and what makes a good reference grammar. Thank you as well to all of the staff at de Gruyter Mouton, in particular to Julie Miess, for making the preparation and publication of this grammar a smooth and enjoyable process.

Finally, I gratefully acknowledge the financial support of the Fulbright Institute for International Education and the National Science Foundation (grant numbers BCS-0537435 (2006-2008) and BCS-0853364 (2009-2012)), who have funded my time spent in the field. I am grateful to the NSF Graduate Student Fellowship Program for allowing me the time in graduate school to prepare this grammar, and to the UCLA Department of Linguistics for helping to support its publication.

Table of contents

Abbreviations

1	1st person
2	2nd person
3	3rd person
ADV	Adverb
AGT	Agentive
AN	Frozen *an-* prefix
ASSOC	Associative
CAUS	Causative
COP	Copula
DD	Discourse definite
DEF	Definite
DIM	Diminutive
DIST	Distal
EMPH	Emphatic
EXIST	Existential
EXP	Experiential
FACT	Factitive
FOC	Focus
FR	French
GER1	Gerundive (-ílé)
GER2	Gerundive (-íyé)
HORT	Hortative
HUM	Human
IMPER	Imperative
IMPF	Imperfective
INF	Infinitive
LOG	Logophoric
MP	Mediopassive
NAA	Prohibitive dummy verb
NEG	Negative
NF	Non-final
NOM	Nominal
OBJ	Object
OBL	Oblique
OPT	Optative
PFV	Perfective
PFV.L	Defocalized perfective
PL	Plural
POSS	Possessive
PPL	Participle
PRO	Pronoun
PROH	Prohibitive
PROX	Proximal
PST	Past
Q	Question
QUOT	Quotative
RECIP	Reciprocal
RED	Reduplicant
REL	Relative
REV	Reversive
SG	Singular
TOP	Topic
⇒	Intonational lengthening
↑	Question intonation

Chapter 1
Introduction

1.1 Dogon languages

The Dogon language family comprises approximately twenty languages spoken on the plains and mountains in the eastern part of Mali's Région de Mopti, with small crossover into neighboring Burkina Faso. The position of Dogon within the larger Niger-Congo macro-family has been the subject of much debate. In particular, the past 70 years have seen it classified as Voltaic (Baumann and Westermann 1948, Bertho 1953, Manessey 1981), Mande (Holas 1951, Delafosse 1952), and Gur (Westermann and Bryan 1952 and Greenberg 1963); however, the current hypothesis (Blench 2005) is that the Dogon family should form its own branch of Niger-Congo. The lack of detailed grammatical descriptions of the Dogon languages has no doubt exacerbated this situation.

At present, the internal classification of the language family also remains to be determined – another consequence of the lack of description. The Dogon Language Project was founded by Dr. Jeffrey Heath of the University of Michigan in an attempt to fill this gap. Since 2004, Dr. Heath and other members of the project including myself have worked to create in-depth grammars of all Dogon languages. In addition, project members have also produced a comparative lexicon of over 8000 entries, which facilitates preliminary grouping of the languages in terms of lexical similarity.[1] As the project and technology progress, we aim to make these data as maximally available as possible for researchers interested in the Dogon language family.

1.1.1 Geographic distribution of the Dogon languages

In the absence of internal genetic classifications, we can group the Dogon languages based on their geographic distribution in so-called Dogon Country.

Northeast	Toro Tegu	(Heath, in preparation;a)
	Bankan Tey	(Heath, in preparation;b)
	Ben Tey	(Heath, in preparation;c)
	Jamsay	(Heath 2008)

1 At the time of writing, the comparative lexicon and all of the unpublished grammar manuscripts listed below are available on our project website at www.dogonlanguages.org.

North central	Nanga	(Heath, in preparation;d)
	Yanda Dom	(Heath, in preparation;e)
	Tebul Ure	(Heath, in preparation;f)
	Ana	(Heath, in preparation;g)
Northwest	Najamba-Kindige (AKA Bondu So)	(Heath, in preparation;h)
	Tiranige Diga (AKA Duleri)	(Heath, in preparation;i)
Central plateau	Tommo So	(current volume)
	Donno So	(Kervran and Prost 1986)
	Bunoge (AKA Korandabo)	(Heath, in preparation;j)
	Dogulu Dom	(Cansler, in preparation)
West-central	Mombo	(Prokhorov, in preparation;a)
	Ampari	(Prokhorov, in preparation;b)
Eastern cliffs	Toro So	(Calame-Griaule 1968; Moran, in preparation)
South-central	Tengu-Kan	(Heath, in preparation;k)
	Togo-Kan	(Heath, in preparation;l)
Southwest	Tomo-Kan	(Dyachkov, in preparation)

One may notice that many of the language names appear to be bipartite or compound in structure. This is because most Dogon language names are made up of the name of the ethnicity or location followed by a word meaning 'language' (*Kan, So, Dom, Tey, Diga*, etc.). There is a certain redundancy in saying "the Tommo So language", but these are the naming conventions based on previous work and on speakers' own preferences, and hence we follow suit.

The references given after each language indicate grammatical descriptions, either published or in progress. Some languages, such as Toro So, represent dialect clusters which may end up being split into multiple languages, pending further investigation.

The list above should not be understood as representing genetic similarity. For example, despite the fact that Tommo So and Dogulu Dom are both spoken in the central areas of Dogon country, the preliminary results of our current fieldwork suggest that Tommo So bears a closer relation in terms of its grammar and lexicon to Najamba or Donno So, while Dogulu Dom bears a closer relation to Mombo and Ampari.

1.2 Tommo So

Tommo So (*Tɔ̀mmɔ̀ Sɔ̀ɔ́*, literally 'Tommo language') is a Central Dogon language spoken on the plateau between Douentza and Bandiagara by an estimated 40,000–

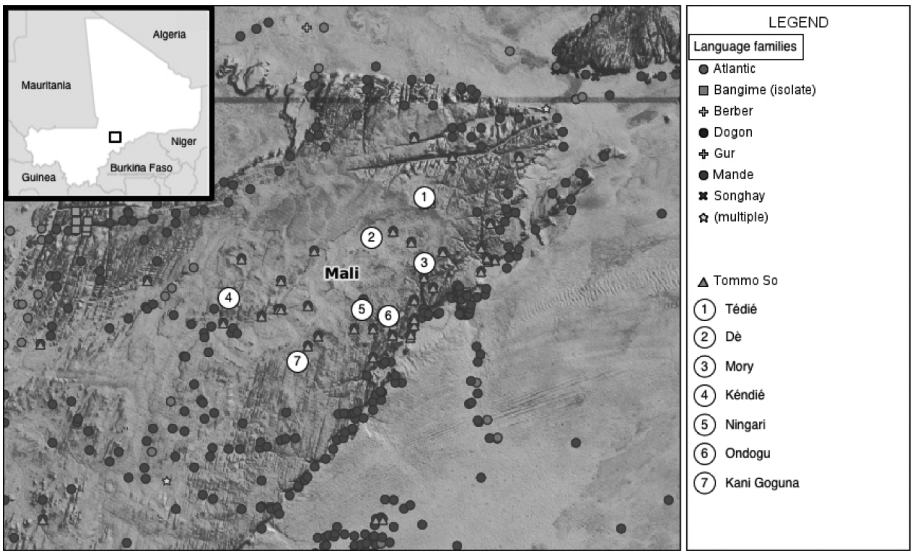


Figure 1: Main Tommo communes on the Bandiagara escarpment

60,000 people (Hochstetler et al. 2004). This makes it the second-most populous language, the most populous being Jamsay with 130,000 speakers according to the same survey. It is considered by the Dogon people to be one of the traditional core Dogon languages. An indication of this is the fact that most songs are sung in Tommo So regardless of what language the singers themselves speak (Hochstetler et al. 2004). The communes[2] of Ningari, Mory, Tédié, Dè, Ondogu, Kani Goguna and Kéndié constitute the heart of Tommo territory; each of these communes has a market whose lingua franca is Tommo So. Figure 1 shows a close-up of the Tommo area. This grammar is based on the northernmost Tédié dialect, which according to recent surveying is considered by Tommo speakers to be the purest and most eloquent form of the language (Hochstetler, p.c.). In terms of neighboring languages, Tommo So is bordered to the northwest by Najamba-Kindige, to the east by Nanga and Jamsay, to the west by Tiranige Diga (known in Tommo So as *Dùlɔ̀ Sɔ̀ɔ́*), and to the south by Donno So and Dogulu Dom.

Based on the available description of Donno So (personal fieldnotes, Kervran and Prost 1986, Farquharson, field notes), this language seems to be Tommo So's closest relative. Donno So is spoken in and around the major market town of Bandiagara, and this geographic location makes it a prominent language in the

2 Mali is divided into *régions*, which are subdivided into *cercles*, and finally into *communes*. Communes tend to be made up of several villages, each of which has a traditional chief. Thus, most of my work was done in the village of Tongo-Tongo, part of the commune of Tédié, in the circle of Douentza and the region of Mopti.

area. Until recently, Tommo So was not listed as a separate language in the Ethnologue (Lewis 2009), simply due to the fact that Donno So was documented earlier by Catholic missionaries (Kervran and Prost 1986) and the two languages were deemed close enough to be considered dialects. While it is true that some dialects of the two languages are mutually intelligible, particularly those dialects of Tommo So closest to Bandiagara (e.g. Kani Gogouna), the data I have seen suggest that Donno So is an intermediate step between two poles formed by Tommo So and Toro So, with the nominal domain closely resembling Tommo So and the verbal domain closely resembling Toro So.

1.3 Environment

Tommo villages are all situated on the plateau of a rocky inselberg[3] mountain known as the Bandiagara Escarpment, located between Douentza and Bandiagara; none can be reached by paved roads. The rocky paths are most easily navigable by motorcycle or on foot, though a well-built 4×4 can get onto the plateau via the gravel roads at Dogani or Bandiagara.

The landscape consists of areas with large boulders interspersed with pebbly or clayey plains and punctuated by small densely wooded copses. The climate is arid, with a rainy season between June and September, when millet (*Pennisetum glaucum*) is cultivated along with other minor crops such as sorghum (*Sorghum bicolor*), cow-pea (*Vigna unguiculata*), sesame (*Sesamum indicum*), roselle (*Hibiscus sabdariffa* in red and green varieties), okra (*Hibiscus esculentus*), peanut (*Arachis hypogaea*), and groundnut (*Vigna subterranea*). Wild fruits such as shea fruit (*Vitellaria paradoxa*), wild grapes (*Lannea microcarpia*), and *zaba* (*Saba senegalensis*) are also harvested at various points throughout the year.

Aside from farming, Tommo people also do a small amount of herding, mainly of sheep, goats and cows. Most of the herding knowledge appears to have originated with the Fulani people, indicated by the near total lack of native Dogon words for practices of animal husbandry. Large game animals have mostly disappeared from the region, but small mammals, such as hedgehogs, mongooses, and rock dassies, as well as many species of insects, birds, and reptiles, still thrive.

In the dry season, there are few ponds or other natural bodies of water; wells and pumps provide water to most villages in the region. Given this arid climate, plant life is confined for most of the year to those trees and shrubs not requiring much water, such as the doum palm, indigo, baobab, and a few species of *Acacia* trees. When the rains arrive, the sandy soil bursts to life with many varieties of grasses, and *Ipomoeia* flowers line the banks of nascent streams.

3 Inselbergs, literally “island mountains”, are small isolated mountains that rise abruptly from level surroundings.

Figure 2: A pile of newly chiseled stone bricks in Tongo-Tongo, Mali

The Dogon people are famous for their mud brick architecture, and Tommo villages are no exception. In the commune of Tédié, stone bricks are hammered and chiseled into shape from huge boulders (see Figure 2). Once the walls of a house have been constructed out of these stone bricks and the roof from tree trunks and branches, the whole structure is plastered with mud brick. Houses are clustered close together in villages, which are surrounded by the residents' fields.

1.4 Culture

The traditional religion of the Dogon people was animist, though this has been rapidly disappearing with the arrival of Islam and Christianity in the area. In the village I worked in, most people identify as Muslim and attend one of the area's mosques on religious holidays. However, some older people maintain animist traditions and the village still contains a mud fetish; as an outsider, it is hard to know how deeply these animist traditions run.

The family structure is patriarchal, and men will often have more than one wife. Families typically live in compounds with small houses and granaries opening onto an inner courtyard, where domestic animals like goats and chickens may be kept.

In addition to Islamic holidays like Ramadan and Eid al-Adha (known locally as Tabaski), the village celebrates traditional Dogon festivals. One of the biggest festivals of the year is Ondom Piri, wherein a series of dances take place over several days. People compose songs for these dances to point out the wrongdoings of community members in the previous year as a way of social policing.

1.5 Language use and vitality

My experience in Mali suggests that Tommo So is not immediately threatened. Children in villages still learn it as their first language and use it almost exclusively until they are school-aged, when they begin to learn French. Even then, the usage of French is limited to the classroom or to interactions with foreigners such as myself. Bilingualism in another Dogon language, typically Jamsay or Najamba, is not uncommon but is by no means universal. Men in particular may also speak Fulfulde, with this language typically acquired through herding experience. People who have spent time in other parts of Mali may also speak Bambara, the lingua franca of the country and the language of most radio broadcasting.

1.6 Tommo So sources

1.6.1 Previous work

To my knowledge, the only scholarly work published on the grammar of Tommo So is *Dogon* (1995), written by Vladimir Plungian for LINCOM's Languages of the World series. While providing a good introduction to the language, the sketch is brief and the author does not mark tone, a key grammatical feature of all Dogon languages. Further, his work is based on the dialect spoken around Ningari, which differs slightly from the dialect discussed here (see Chapter 22). Several articles have appeared over the last fifty years that mention Tommo So (often under the name Tombo So) among other dialects of Dogon, including several by Plungian. For a detailed bibliography, see Hantgan (2007).

In terms of lexical work, the Direction Nationale de l'Alphabétisation Fonctionnelle et de la Linguistique Appliquée (D.N.A.F.L.A., now known as the Institut National des Langues – Abdoulaye Barry) has published *Eléments de Terminologie Dogon* (1984) based on Tommo So, and several other articles provide limited wordlists (Arnaud 1921, Barth 1912, Bertho 1953, DNFLA/DRLP 1981, Galtier 1993, Ongoiba 1988).

1.6.2 Current fieldwork

The data in this grammar are from work with speakers from the commune of Tédié (specifically the villages of Tongo-Tongo and Anji) from June 2008 to February 2012. Elicitation was done both *in situ* and in the nearby towns of Douentza and Sévaré with consultants from Tédié. Most elicited data were provided by two speakers, a young woman from Tongo-Tongo (Ramata Ouologuèm) and a young man from Anji (Issa Toloba). Elicited data were supplemented by naturalistic data from recorded texts. Audio recordings were made using a Zoom H4 digital recorder and analyzed using Praat (Boersma and Weenink 2011). Video recordings were made using a Sony HDR-CX190 digital HD camcorder. In addition to notebooks, data were stored in Microsoft Word, Excel, and Filemaker Pro.

Chapter 2

Grammatical sketch

This chapter will provide an introduction to the main grammatical features of Tommo So, all of which will be covered more in depth later in the grammar.

2.1 Phonology

2.1.1 Segmental inventory and phonotactics

The phonemic inventory of Tommo So consists of 17 consonants and 7 vowels, for which length and nasalization are contrastive. ATR, backness, and height harmony is present in stems and to a lesser extent between stems and derivational affixes; almost all inflectional affixes are outside the domain of harmony. An interesting feature of Tommo So harmony is that it is variable, and the rates of harmony correlate with morphological distance from the stem. For further discussion, see section 3.5.

Unlike some Dogon languages, /l/ and /r/ are contrastive, but neither liquid occurs in word-initial position. Only sonorant consonants can serve as codas.

The transcription system used in this grammar is a modified form of IPA. The divergences are intended to make it closer to standard orthographies in use for other West African languages like Bambara. Thus, affricate /ʤ/ is written 'j', tap /ɾ/ as 'r', and /j/ as 'y', while long vowels are written as two consecutive vowels ('aa' rather than /a:/). Nasalization is marked with a superscript n {vⁿ} to distinguish it from nasal codas; this also ensures maximum visibility of tone marking, since the use of tilde would require stacking diacritics.

2.1.2 Tonal inventory and tonotactics

Like all Dogon languages, Tommo So is tonal, with tonal primitives H[igh] and L[ow], marked {v́} and {v̀}, respectively. It also has two contour tones, <LH> rising ({v̌} or {v̀v́}) and <HL> falling ({v̂} or {v́v̀}), though instances of the latter are rare. Tommo So lacks the "bell-shaped tone" (following Heath 2008) <LHL> characteristic of other Dogon languages such as Jamsay or Nanga, but it does maintain a three-way contrast between H, L and toneless, with the latter constrained to clitics and certain suffixes (McPherson 2011).

Native Tommo So stems all have one of two melodies: /H/ or /LH/, but loanwords have introduced /HL/ and /LHL/ tone melodies into the lexicon. /HLH/ is not a licit surface melody. Word-level tone melodies can be overwritten in certain grammatical contexts; these changes are listed at the end of this sub-section and described in depth in Chapter 4.

2.1.3 Key phonological alternations

One salient feature of Tommo So phonology is the variable epenthesis of [u] after sonorants. While Plungian (1995) lists 'sun' as *nàmú*, most of the speakers in Tédié will say simply *năm*. On other words, like 'white', they may vary, pronouncing either *pílu* or *píl*. The fact that the vowel is epenthetic is clear from the fact that it is toneless and does not harmonize with the stem. This will be discussed further in section 3.4.6.

Vowel hiatus is pervasive in the morphophonology. Many suffixes are vowel-initial, and most stems are vowel-final. When these two vowels come together, the final vowel of the stem deletes:

(1) *káná-ee → kán-ee* 'doing (non-final)'

See section 3.7.4.

A final phonological effect to note is that like other Dogon languages such as Jamsay (Heath 2008: section 3.5.3), Tommo So shows the effects of a single left-aligned trochee, in that the second syllable, particularly in longer words, tends to be metrically weak. This results in both vowel syncope and vowel reduction. For examples and discussion, see section 3.6.

2.1.4 Key tonal changes

Grammatical tone is very prevalent in Tommo So, with typically word level tone changes in both the nominal and verbal realms. The following summarizes some possible tonal overlays:

Overlaid all L ("tone lowering")
NP as head of a relative clause
Relative participle before demonstrative
NP before adjective or demonstrative
Possessed NP following full NP (non-pronominal) possessor
First stem in canonical and synthetic compounds
Second stem in pseudogenitive compounds
Verb stem before negative suffixes

Overlaid all H
Singular affirmative imperative of most verbs
Inalienably possessed nouns (1–2 moras) with a pronominal possessor
Verbal stem in gerundive compounds
Certain nominalized verbs

Overlaid HL
Perfective and negative relative participles
Possessed kinship terms (3+ moras) with a pronominal possessor
Verb stems in the imperfective affirmative

Overlaid LH
Certain nominalized verbs

For more on grammatical tone, see section 4.4–5.

2.2 Verbal inflection

The basic word order is SOV; the inflected verb (be that auxiliary or otherwise) comes at the end of the sentence. Like nominal stems, verb stems in Tommo So belong to one of two tonal classes: /H/ or /LH/ with the L on the first syllable or mora (on monosyllabic verbs). In the rare monomoraic verbs, /LH/ surfaces as simply [L] in the absence of a second mora. Depending on the inflection, the H may surface on the following suffix. Which tone pattern a verb stem will take is partially predictable by the first consonant, with voiced obstruent-initial stems taking /LH/, voiceless obstruent- and vowel-initial stems taking /H/, and sonorant-initial stems lexically listed.

As indicated in section 2.1.4, the lexical tone of the stem is often overridden by grammatical tone patterns when inflected. I will discuss this in conjunction with verbal inflectional paradigms below.

The basic form of the verb in main clauses is as follows:

(2) [stem- (derivational suffix(es))] – aspect/negation – subject marking

This final subject marking is achieved through the addition of the following subject suffixes:

(3)

1sg	*-m*	1pl	*-y*
2sg	*-w*	2pl	*-y*
3sg		3pl	(-N)

I have placed the third person plural suffix in parentheses because there is no single suffix that surfaces in each inflected form. Rather, the third person plural seems to have at least some nasal element that morphs and fuses with the aspectual suffix.

Verbal inflection is marked in two ways: by discrete inflectional suffixes and by stem-level tone overlays; they typically cannot be divorced one from the other. For the full treatment of aspect and verbal inflection, see Chapter 12.

2.2.1 Aspect

There are two affirmative perfective forms for each verb, one of which carries the suffix *-aa* and the other of which carries one of three allomorphs [-i, -e, -ɛ]; I abbreviate this latter form as the E-perfective. The choice of allomorph is largely predictable by stem length and vowel harmony, with stems with three or more moras always taking the allomorph [-i], but some stems are lexically listed for one allomorph or the other. See section 12.4.

The two perfective forms differ in their tonal contours as well. The *-aa* form has no effect on the lexical tone of the stem. The second perfective form, however, does change the tone of the stem. If the verb is defocalized, it undergoes tone lowering (its tone is overwritten with {L}). If it is focused and reduplicated, it takes {HL} with a L-toned initial reduplicant.

The choice of which perfective form to use is dependent on focus and relativization. The E-perfective is used in relative clauses and usually when there is a nonverbal focused element in the phrase, and the AA-perfective is used elsewhere. A reduplicated form the E-perfective is used when the verb itself is focused.

The imperfective is used for the habitual and the future, and it takes the suffix *-dɛ̀* (*-dɛ̀ŋ* for 3pl). In this form, the verb takes a {HL} overlay with H on the first mora.

The progressive is made up of a participle and an auxiliary verb, either 'have' *=sɛ* or 'be' *=wɔ*. In my consultants' speech, there is no discernible difference between the two forms. In the dialect described by Plungian (1995), however, the former is used for the simple progressive and while the latter is used for an iterative progressive (Plungian 1995). The participial suffix is *-gú* (occasionally *-nú*), which comes with no grammatical tone overlay. Because of this, I typically use the progressive to identify the underlying stem. Stems listed in the lexicon are based off of the progressive.

2.2.2 Negation

The basic form of the negative suffix is /-lV/, with the vocalism of 'V' depending on the aspect of the inflected verb. In all negative forms, the stem takes a {L} overlay. The distinctions seen in the affirmative perfective are collapsed into one negative form. Here, the negative takes the suffix *-lí* (*-nní* for 3pl.), suffixed onto the verb stem. For the imperfective, the habitual/future negative takes the suffix *-éélɛ̀* (*-énnɛ̀* for 3pl). The negative progressive merely inflects the auxiliary, which we will come to in a more detailed discussion of auxiliary and modal verbs (see Chapter 13).

2.2.3 Imperative and hortative

The imperative makes a distinction between second person singular and plural, with the singular having no overt marking and the plural the suffix *-ɲ̀*. The imperative stem variably takes either a {H} overlay or no overlay. A falling tone is formed on

the last syllable of the plural imperative due to the presence of the L-toned suffix. The negative imperative has the suffix *-gú*, which becomes [-gîŋ] in the plural from the vowel fronting in the combination of *-gú* + *-ǹ*.

The hortative is interesting in that here a distinction is made between first person dual and plural. The dual (you and I) suffix is *-mɔ́*, while the plural (you all and I) is *-mɔ́-ǹ*. Once again, this can be analyzed as the plural suffix from the imperatives, and so these hortatives are better seen as meaning "me and one single person" and "me and several people", respectively.

2.3 Verbal derivation

There are five main derivational suffixes in Tommo So: factitive *-ndɛ*, transitive *-irɛ*, reversive *-ilɛ*, mediopassive *-iyɛ*, and causative *-mɔ*. When more than one suffix is present, they surface in roughly this order. As mentioned in section 2.1.1 above, one of the most interesting aspects of Tommo So morphophonology is the variable application of vowel harmony in these derivational suffixes. For more on verbal derivation, see Chapter 11.

2.4 Noun phrase (NP)

The NP can be maximally made up of the following constituents, in order:

(4) a. possessor NP and/or pronoun possessor
b. noun stem
c. adjective(s)
d. (numeral)
e. possessive pronoun
f. demonstrative, definite
g. plural particle
h. (numeral)
i. 'all'

There is some variation in the placement of the numeral, shown by its double listing in parentheses above. For a discussion of what licenses this variation, see section 7.3. An example of a fairly complex NP is given below:

(5) *yàà-nà*^{L} *èsú=gɛ* *jàndùlù*^{L} *gɛ̀m*^{L}*=gɛ=mbe* *tààndú-go*
female-HUM.SG pretty=DEF donkey black=DEF=PL three-ADV
'the pretty woman's three black donkeys'

The NP is the site of many interesting tonal interactions, since both demonstratives and adjectives force tone lowering on the preceding noun, be it simple or compound, while numerals do not interact tonally at all. Nonpronominal NP possessors cause tone lowering on the following noun. Tommo So makes an inalienable/alienable distinction in possession, which is evident both from differing domains of grammatical tone and the use of different possessive pronouns for each type of possession. Inalienable nouns consist of kinship terms and practically nothing else. For more on the NP, see Chapter 7.

2.5 Case marking and PPs

The object can be optionally marked for case with the enclitic *=ɲ̀*. Human objects, both direct and indirect, are obligatorily case marked if only one human object is present; if two are present, the first of the two objects may be bare. For more discussion, see section 13.1.6. Subjects are not case-marked.

Tommo So also has postpositions, which are grammatically enclitics: oblique *=nɛ* meaning 'in', 'on' and occasionally 'for'; associative *=le* meaning 'with' or 'and'; locative *=baa* meaning 'at'; possessive *=mɔ* meaning 'for' or 'belonging to'. Indirect objects like benefactives may be marked with postpositions. The distinction between 'in' and 'on' is made by both context and the choice of copula. More complex spatial relations are made by adding a postposition (typically *=nɛ* or *=baa*) after a body part word such as *ónnu* 'back' or *gíré* 'face', which is then used in a genitive construction with object of reference for location. For example, to say that the child is behind the chair, one would say 'the child is at the chair's back'. See Chapter 10 for more on postpositions.

2.6 Main clauses and constituent order

As previously stated, the basic word order of main clauses in Tommo So, like many Dogon languages, is SOV. Temporal adverbials such as *yògó* 'tomorrow' are generally clause-initial, followed by the full-NP subject (if there is one), indirect object or direct object (variable order), and the verb, loosely in that order. Pronominal subjects are expressed through verbal suffixes; independent subject pronouns, which come clause-initially, are generally only used if the subject is focalized, as illustrated in (6).

(6) a. *Mòòmíyó mí=ɲ̀ támbá-gú=sɛ.*
scorpion 1SG.PRO=OBJ strike-PPL=have
'The scorpion is striking me.'

b. *Mí áí^{H}=ɲ̀ màŋgóró ób-aa=be-m.*
1SG.PRO friend=OBJ mango give-PFV=be.PST-1SG
'I gave my friend a mango.'

c. *Yògó bɔ̀gɔ̀^{L} èsú úwɔ ɛ́bɛ̀-dɛ̀-m.*
tomorrow dress pretty 2SG.POSS buy-IMPF-1SG
'Tomorrow I will buy you a pretty dress.'

d. *É émmé=le Tɔ̃mmɔ̀^{L} Sɔ̃ɔ́ sɔ́ɔ̀-dɛ̀-y.*
2PL.PRO 1PL.PRO=ASSOC Tommo speech speak-IMPF-2PL
'*You* speak with us in Tommo So.'

Note that in example (6b), the independent pronoun *mí* is acting as a possessor, not as a subject.

2.7 Relative clauses

In relative clauses, the relative participle is devoid of subject inflection. In Jamsay, it is treated nominally, evident from the fact that it agrees with the head noun in nominal features like animacy, but this marking is very limited in Tommo So as a whole and is not present on the relative participle. Relative clauses are typically head-internal, and the head takes the all {L} overlay typical of other modifiers. The subject of the relative participle, if pronominal, is expressed by an independent pronoun before the verb. If the relative clause is modified by a demonstrative, then the tone of the participle is lowered as well, but not the tone of its subject or any other non-head constituents. If the head of the relative clause is a possessor plus possessed noun combination, tone lowering does not take effect; that is, possessives are impervious to tone lowering beyond what is already imposed by the possessor.

(7) a. *Aràmátá nàà^{L} gɛ̀m^{L} mí sɛ́m-ɛ̀=gɛ*
Ramata cow black 1SG.PRO slaughter-PFV.REL=DEF
yúú ḿmɔ tɛ́mɛ́-gú=be.
millet 1SG.POSS eat-PPL=be.PST
'Ramata's black cow that I slaughtered used to eat my millet.'

b. *Àn-nà^{L} sɔ̃ɔ́ sɔ̃ɔ̀-gù^{L} mí sɛ̀^{L} nɔ́*
man-HUM.SG speech speak-PPL 1SG.PRO have this
mí báá^{H} íg-go=wɔ.
1SG.PRO father know-ADV=be
'This man I am speaking to knows my father.'

2.8 Interclausal syntax

There are a number of ways to combine clauses. I have schematically listed a few of the more common ways below to be discussed in greater depth in Chapters 17–19.

(8)

	Structure	Typical function
a.	Chaining	
	...verb suffixed with *-ee*	Same subject VPs, imperfective
	...verb suffixed with *-aa*	Same subject VPs, perfective
b.	Conditionals	
	[...inflected verb]=*yo*	If (when)...
c.	Adverbials	
	[...participle]	Temporal, by... (while...)
	[...bare verb stem]=*nɛ*	Before...
d.	Quotative	
	[...inflected verb] 'say'	Quotative (reported assertion)
e.	Complement	
	[...bare verb stem]	Complement (with 'want (past)', 'begin')
	[...inflected verb]	Complement (with 'finish', 'can' etc.)
	[...infinitive]	Complement (with 'want', 'be afraid', 'forget to', etc.)
f.	Purposive	
	[...participle]	Purposive (with 'go', 'come', etc.)

In each of the above, it is understood that a main clause follows what is found in the first column. For instance, an illustration of (8a) would be:

(9) *Pédu=gɛ* *píy-aa* *kúú=gɛ* *jìjìb-ì.*
sheep=DEF cry-PFV head=DEF shake-PFV.L
'The sheep cried and shook its head.'

Chapter 3

Segmental phonology

This chapter deals exclusively with segmental phonology: phoneme inventory, syllable structure, phonotactics, vowel harmony, and phonological rules. For a description of tonal phonology, see the next chapter. All minimal pairs, frequency counts, and phonotactic generalizations are based on a lexical corpus of 8930 entries. Further data on vowel harmony (section 3.5) are drawn from field notes.

3.1 Consonants

3.1.1 Consonant phoneme inventory

Seventeen consonant sounds can be considered phonemic in Tommo So based on their appearance in regular (non-ideophonic) lexical items and their involvement in minimal pairs; these are summarized in the following chart. With the exception of [h], all other phonemes are native to Tommo So. Here and throughout the grammar, Tommo So will be written in a practical orthography inspired by the standard orthographies of local languages like Bambara but differing in a few regards (i.e. the use of superscript rather than plain 'n' to mark nasalization). In the chart below, IPA is given in brackets following the letter of the practical orthography where the two differ:

(10) *Tommo So consonant inventory*

	Bilabial	Alveolar	Alveolo-palatal	Palatal	Velar	Glottal
Plosive	p b	t d			k g	
Nasal	m	n		ɲ	ŋ	
Fricative		s				h
Affricate			j [dʑ]			
Approximant	w			y [j]		
Tap		r [ɾ]				
Lateral approximant		l				

3.1.2 Exceptional sounds

In addition to the phonemes listed above, there are a few consonant sounds that are not in variation with (allophones of) any phoneme. However, they are found in such a limited range of examples that I choose not to consider them part of the phoneme inventory. The first of these is the voiced alveolopalatal fricative [ʑ], which is attested only in the word *ʑìmààgú* '*Sarcostema viminale* (plant species)'. While this sound

would help fill out the alveolo-palatal place of articulation, otherwise represented only by the affricate /dʑ/, it is difficult to argue for the phonemic status of a sound based on one rather obscure example alone. For this reason, I leave [ʑ] out of the phoneme chart pending further data.

Similarly, ideophones (see Chapter 8) often contain consonants that are not part of the regular phoneme inventory, a common phenomenon cross-linguistically. One such sound found in ideophones but not in regular lexical items is glottal stop [ʔ] (practical orthography [']), as in *ku' ku' ku'* (sound of two women pounding millet). Another example is the voiceless alveolo-palatal affricate [tɕ] (practical orthography [c]), found in ideophones like *cákàm-cákàm* (noisy chewing). Given that neither consonant is found in regular vocabulary, I do not consider them to be phonemes of the language.

3.1.3 Gemination

Consonant length is phonemic, but in native words, only sonorants are geminate. At first glance, this seems puzzling from a typological perspective, particularly the presence of typologically rare geminate glides and not typologically common geminate stops. However, this phenomenon arises from the fact that the only permissible coda consonants in Tommo So are sonorants; an intervocalic geminate stop would result in an ill-formed syllable.

The distribution of sonorant geminates is further restricted in stems to nasals and one instance of geminate /ww/. Geminates [ll] and [yy] can arise from the morphophonology, but geminate [rr] is not attested (in native words). The geminate velar nasal /ŋŋ/ is also not attested, despite the presence of other geminate nasals. The following minimal or near-minimal pairs demonstrate the length distinction in consonants. Where gemination is the result of a morphophonological rule, I include the base form.

(11)	a.	*dɛ̀nɛ́*	'spend the day'	*dɛ̀nnɛ́*	'look for'
	b.	*dàmá*	'taboo'	*dàmmá*	'type of hoe'
	c.	*gɪ̀ŋɛ́*	'beg'	*gɪ̀ŋŋɛ́*	'intersperse'
	d.	*úwɔ*	'yours'	*túwwɔ́*	'nine'
	e.	*dùyɔ́*	'insult'	*dùyyɔ́*	'carry' < *dùy-íyɔ́*
	f.	*yɛ̀lɛ̀*	'he came'	*yɛ́llɛ̀*	'he'll come' < *yɛ́lɛ̀-dɛ̀*

The geminate approximants [ll] and [yy] derive from vowel syncope, and in the case of [ll], subsequent regressive assimilation of /d/ to [l]. See section 3.6.2 for further discussion of this rule.

The phonemic status of geminate /ww/ is tenuous at best. In addition to the fact that it is attested in only one word, *túwwɔ́* 'nine', singleton /w/ in this context (u_ɔ)

is only attested in the 2sg possessive pronoun *úwɔ*, itself historically derived from the independent pronoun *ú* and a possessive clitic *=mɔ*. It is unsurprising that /w/ in this context would be rare, given the perceptual difficulties involved in hearing a labial glide between two round vowels. It is possible that *túwwɔ́* 'nine' was historically a singleton that strengthened to a geminate in order to remain distinct from its neighboring vowels. This tentative explanation is bolstered by the form of the numeral 'nine' in other Dogon languages, such as Jamsay *lá:rúwà* (Heath 2008) and Najamba *twây* (Heath, in preparation;h), each of which contain a singleton /w/.

In (11c), a minimal pair is given with singleton and geminate /ɲ/, but in fact only the geminate surfaces as a nasal stop. Intervocalic singleton /ɲ/ in Tommo So has weakened into a nasalized palatal sonorant [yⁿ], with words like *gìŋɛ́* 'beg' pronounced as [gìyⁿɛ́]. This development has parallels in the diachrony of other Dogon languages, such as Jamsay, where singleton **m* became the phoneme /wⁿ/ while historical geminate **mm* has become a singleton /m/. Thus we see cognates like Jamsay *nàmá* and Tommo So *nɔ̀mmɔ́* 'Dogon water god', where a geminate in Tommo So corresponds to a singleton in Jamsay, and pairs like Jamsay *nàwⁿá* and Tommo So *nàmá* 'meat', where a Tommo So singleton nasal stop corresponds with a Jamsay nasalized approximant. In Tommo So, the same shift is taking place at the palatal place of articulation, but degemination has not yet occurred, and as such we have no evidence for treating [yⁿ] as anything but an allophone of the palatal nasal. See section 3.1.4.9 for more on /ɲ/.

Though native Tommo So words do not allow geminate obstruents, for the phonotactic reasons pointed out above, we do find such geminates in loanwords, particularly from Fulfulde. Examples include *sɔ́bbɔ̀* 'dry sowing' or *éddɛ̀* 'widow's four month period of mourning'. However, these geminates are relatively rare. See section 3.4 on phonotactics for token numbers in the lexicon.

3.1.4 Minimal pairs

The following subsections address each phoneme in detail, showing allophonic variation where applicable and providing minimal and near minimal pairs with similar phonemes to illustrate their phonemic status.

3.1.4.1 /p/

Voiceless stops /p, t, k/ are all unaspirated in Tommo So, with an average VOT (voice onset time) of around 20ms. Since the Dogon languages lack a labiodental fricative /f/, it is characteristic of Dogon speech to replace all initial /f/ in loanwords with /p/. Thus, the Fulani people are referred to as *púlɔ̀*, and a French loanword like *fête* 'party' will often be pronounced *pɛ́du*. Due to the fact that voiceless stops only occur

in word-initial position, medial /f/ will typically co-vary with /b/ in this position rather than /p/, as in *màlfá ~ màlbá* 'gun'. See section 3.4 for more on phonotactic restrictions.

The following pairs show that voicing is contrastive for labial stops (/p/ vs. /b/), that place of articulation is contrastive for voiceless stops (/p/ vs. /t, k/) and that nasality is contrastive for labial stops (/p/ vs. /m/).

(12)

	Minimal pairs for /p/ vs...			
/b/	*pílu*	'white'	*bílu*	'ladder'
/t/	*pɛ́ɛ́ⁿ*	'harden'	*tɛ́ɛ́ⁿ*	'hobble'
/k/	*pɛ́dɛ́*	'gin (cotton)'	*kɛ́dɛ́*	'cut'
/m/	*pàndé*	'widowhood'	*màndé*	'Mande'

3.1.4.2 /b/

The last subsection showed that /b/, fully voiced in Tommo So, often stands in for a word-medial /f/ in loanwords. It is also stands in for /v/ in French loanwords like *avion* 'airplane', yielding Tommo So pronunciation *àbíɔ̀ⁿ*. Native /b/ undergoes no allophonic variation.

Voicing, nasality, place, and sonorancy are all contrastive for voiced stops.

(13)

	Minimal pairs for /b/ vs...			
/p/	*bílu*	'ladder'	*pílu*	'white'
/m/	*bɔ̀ɔ̀ndɔ́*	'pamper'	*mɔ̀ɔ̀ndɔ́*	'stock up'
/d/	*bàbá*	'respect'	*bàdá*	'heal'
/g/	*bòó*	'call'	*gòó*	'go out'
/w/	*jàbá*	'replaster'	*jàwá*	'branch out'

3.1.4.3 /t/

Like /p/, voiceless alveolar /t/ is also unaspirated. The following minimal pairs demonstrate the phonemic status of the voiceless alveolar stop with regards to voiced alveolar stops, nasal alveolar stops, voiceless stops at other places of articulation, and finally frication, showing that /t/ contrasts with /s/.

(14)

	Minimal pairs for /t/ vs...			
/d/	*tɛ̀bɛ́*	'cross pole'	*dɛ̀bɛ́*	'get stuck'
/n/	*táá*	'shoot'	*náá*	'mother'
/p/	*tɛ́ɛ́ⁿ*	'hobble'	*pɛ́ɛ́ⁿ*	'harden'
/k/	*tɛ́mɛ́*	'eat (meat)'	*kɛ́mɛ́*	'pinch'
/s/	*tɛ́mɛ́*	'eat (meat)'	*sɛ́mɛ́*	'slaughter'

3.1.4.4 /d/

Before high front /i/, /d/ becomes slightly affricated. While this means that its pronunciation approaches that of /j/, the two sounds remain distinct, as indicated by the presence of minimal pairs. Likewise, /g/ and /d/ also contrast before /i/, showing that the palatalized allophones of each are not close enough to result in neutralization. See section 3.1.4.6 for a discussion of /g/'s allophonic behavior.

The following pairs show that /d/ is contrastive for voicing with /t/, for nasality with /n/, and for place of articulation with /b/ and /g/.

(15)

	Minimal pairs for /d/ vs...			
/t/	*dɛ̀bɛ́*	'get stuck'	*tɛ̀bɛ́*	'cross pole'
/n/	*dàá*	'kill'	*nàá*	'forget'
/b/	*bàdá*	'heal'	*bàbá*	'respect'
/j/	*dìŋ̂ɛ*	'tie'	*jìŋɛ́*	'approach'
/g/	*dìnɛ́*	'(sb's) turn'	*gìnɛ́*	'house'

3.1.4.5 /k/

The phoneme /k/, also unaspirated, has the allophone [c] (IPA palatal stop, not orthographic affricate) before high front /i/, but this allophony is purely distributional and does not form the basis of any morphophonological alternations; the underlying sequence /ki/ always surfaces as [ci], but this sequence is never created on the surface from morphological operations. We find the same contrasts in terms of voicing and place of articulation that we did for the other voiceless stops, but the contrast for nasality is less clear-cut. This is due to the fact that in native stems, /ŋ/ cannot occur in word-initial position, the only position in which /k/ can appear. One near-minimal pair (differing in tone) is offered here using the Arabic loanword *jákà* 'Islamic tithe' (from Arabic *zakat*, borrowed via Fulfulde). However, since /k/ contrasts with /g/ and /g/ contrasts with /ŋ/, it seems safe to say that /k/ also contrasts with /ŋ/.

(16)

	Minimal pairs for /k/ vs...			
/g/	*kìyé*	'bone'	*gìyé*	'thorn'
/ŋ/	*jákà*	'Islamic tithe'	*jàŋ́á*	'put (on a stand)'
/p/	*kɛ́dɛ́*	'cut'	*pɛ́dɛ́*	'gin (cotton)'
/t/	*kɛ́mɛ́*	'pinch'	*tɛ́mɛ́*	'eat (meat)'

3.1.4.6 /g/

There are four allophones of the voiced velar stop /g/. These are: [ɟ] before /i/, [gʷ] before round vowels, [ɣ] in the contexts /a_a/ and /ɔ_ɔ/, and [g] elsewhere.

Like /k/, /g/ will also become palatalized before /i/, which makes it difficult for non-native ears to distinguish it from /j/. There are no minimal pairs in the data set for /g/ and /j/ before /i/, which suggests neutralization in this context, but the fact that speakers correct me if I pronounce a word with the wrong phoneme (/g/ for /j/ or /j/ for /g/) shows that it is not true neutralization. Before /i/, /g/ surfaces as a mildly affricated [ɟ], while /j/ continues to surface as [ʤ]. This area would reward phonetic analysis. Though /g/ does not palatalize before the other front vowels /e/ and /ɛ/, there are still no minimal pairs contrasting /g/ and /j/ in this context.

Before a round vowel, the voiced velar stop has a rounded allomorph, yielding pronunciations like [gʷu], [gʷo], and [gʷɔ]. Why this rounding is not present or audible on the voiceless velar stop is not clear.

Finally, /g/ lenites to [ɣ] in some metrical contexts when flanked by two /a/ vowels or two /ɔ/ vowels. For a discussion of this lenition, see section 3.6.3.

The minimal pairs below show that /g/'s voicing, orality, and place are phonemic.

(17) Minimal pairs for /g/ vs...

/k/	*gìyé*	'thorn'	*kìyé*	'bone'
/ŋ/	*pégɛ́*	'button'	*péŋɛ́*	'knot'
/b/	*gòó*	'go out'	*bòó*	'call'
/d/	*gɔ̀ɔ́*	'dance'	*dɔ̀ɔ́*	'arrive'
/j/	*gɔ̀mbɔ́*	'open wide'	*jɔ̀mbɔ́*	'peck at'

3.1.4.7 /m/

/m/ is represented by a single allophone [m], though it can assimilate in place of articulation to a following stop; see section 3.7.2. The minimal pairs below demonstrate its phonemic status with regards to nasality, place of articulation, and continuancy. I have given minimal pairs for both singleton /m/ vs. /ɲ/ and geminate /mm/ vs. /ɲɲ/, since the former are further differentiated by the fact that /ɲ/ surfaces as a nasalized approximant. Further, /m/ is only contrastive with /ŋ/ in intervocalic position due to phonotactic restrictions on the velar nasal.

(18) a. Minimal pairs for /m/ vs...

/b/	*mɔ̀ɔ̀ndɔ́*	'stock up'	*bɔ̀ɔ̀ndɔ́*	'pamper'
/w/	*ámá*	'in-law'	*áwá*	'catch'
/n/	*máá*	'dry'	*náá*	'mother'
/ɲ/	*dìmɛ́*	'tamp down'	*dìŋɛ́*	'prop up'
/ŋ/	*ámá*	'in-law'	*áŋá*	'mouth'

b. Minimal pairs for /mm/ vs...

/ɲɲ/	*ímmɛ́*	'inflate'	*íŋŋɛ́*	'lift (a heavy rock)'

3.1.4.8 /n/

Like /m/, the surface form /n/ does not undergo allophonic variation. The following minimal pairs show the phonemic status of nasality and place of articulation for /n/. The minimal pair with /ɲ/ shows the contrast in word-initial position, a rare position for palatal /ɲ/. The two also contrast intervocalically, as in *dìnɛ́* '(sb's) turn' and *dìŋɛ́* 'prop up'.

(19) Minimal pairs for /n/ vs...

/d/	*nàá*	'forget'	*dàá*	'kill'
/m/	*náá*	'mother'	*máá*	'dry'
/ɲ/	*nǎm*	'sun'	*ɲǎm*	'fire'
/ŋ/	*káná*	'do'	*káŋá*	'discuss'

3.1.4.9 /ɲ/

As discussed in section 3.1.3, /ɲ/ has the allophone [y^{n}] (IPA [j̃]) when the nasal is an intervocalic singleton. This variant may also be used in word-final position, though it is in free-variation with the nasal stop. The minimal pairs below contrast /ɲ/ with a voiced oral affricate, nasal stops at other places of articulation, as well as with the non-nasalized palatal sonorant /y/.

(20) Minimal pairs for /ɲ/ vs...

/j/	*ɲǎm*	'fire'	*jǎm*	'caste of leatherworkers'
/m/	*íŋŋɛ́*	'lift'	*ímmɛ́*	'inflate'
/n/	*ɲǎm*	'fire'	*nǎm*	'sun'
/ŋ/	*dìŋɛ́*	'burn'	*dìŋɛ́*	'tie'
/y/	*ɲám*	'difficult'	*yám*	'wasted, broken'
	[kúy^{n}ɔ́]	'squirrel'	*kúyɔ́*	'first'

The minimal pair *ɲǎm* 'fire' and *jǎm* 'caste of leatherworkers' is in fact the only minimal pair for /ɲ/ and /j/ since /j/ never occurs intervocalically or word-finally in native words.

3.1.4.10 /ŋ/

Before /i/, the velar nasal /ŋ/ is produced at a place of articulation intermediate between velar and palatal. Thus, the nasal in *níŋ-íyɛ́* 'be afraid' is not a canonical velar nasal and approaches /ɲ/ in its pronunciation, but we know the underlying form given other forms of the verb such as *níŋɛ́-ndɛ́* 'scare' with the factitive suffix.

Since the velar nasal cannot appear in word-initial or word-final position, all minimal pairs reflect word-medial contrasts for nasality and place of articulation. See section 3.4 for more on phonotactics.

(21) Minimal pairs for /ŋ/ vs...

/g/	*pɛ́ŋɛ́*	'knot'	*pɛ́gɛ́*	'button'
/m/	*áŋá*	'mouth'	*ámá*	'in-law'
/n/	*káŋá*	'discuss'	*káná*	'do'
/ɲ/	*dìŋɛ́*	'tie'	*dìɲɛ́*	'prop up'

3.1.4.11 /s/

Before /i/, the alveolar /s/ has an optional allophone [ɕ], an alveolo-palatal fricative. For example, the verb *sírɛ́* 'cook' can be pronounced [ɕírɛ́], but in careful speech, [sírɛ́] is also possible. The following minimal pairs show that /s/ contrasts with /t/ in continuancy and with /h/ in place of articulation.

(22) Minimal pairs for /s/ vs...

/t/	*sárá*	'pay'	*tárá*	'stick on'
/h/	*sárá*	'pay'	*hárá*	'be forbidden'

3.1.4.12 /h/

The voiceless glottal fricative /h/ is not a native phoneme, but it has been well integrated into the system of phonemic contrasts due to a large number of loanwords from Fulfulde (often originally of Arabic origin). It also occurs in exclamations like [ɔ̀hɔ́ɔ̀ⁿ] 'yes; I see' and ideophones like [hèéé], an expressive adverbial referring to extreme height. The following minimal pairs contrast /h/ with null to show that word-initial glottal frication is phonemic, as well as with /s/ to show that place of articulation is phonemic.

(23) Minimal pairs for /h/ vs...

Ø	*hárá*	'be forbidden'	*árá*	'suck'
/s/	*hárá*	'be forbidden'	*sárá*	'pay'

3.1.4.13 /j/

The place of articulation of the affricate written /j/ in the practical orthography is alveolo-palatal (IPA /dʑ/), like the voiced affricate in Japanese *jibun* 'self', rather than the English-like post-alveolar affricate (IPA /dʒ/). Though, as section 3.1.4.6 discussed, there are no minimal pairs between /g/ and /j/ before /i/, plenty of minimal pairs are to be found in other environments and the contrast before /i/ does not seem to be fully neutralized. /j/ itself undergoes no allophonic variation.

The pairs below show that /j/ contrasts with /g/ in contexts other than before /i/, /d/ in all contexts (/i/ given here), and with the palatal sonorant /y/.

(24) Minimal pairs for /j/ vs...

/g/	*jàmbá*	'chip off'	*gàmbá*	'reduce'
/d/	*jìŋɛ́*	'approach'	*dìŋɛ́*	'tie'
/y/	*jàbá*	'replaster'	*yàbá*	'agree'

3.1.4.14 /w/

The phoneme /w/ is a labio-velar approximant, like English /w/, and it does not show allophonic variation. The following minimal pairs show the phonemic status of continuancy, nasality, and backness for /w/.

(25) Minimal pairs for /w/ vs...

/b/	*dàwá*	'cover up'	*dàbá*	'hoe'
/m/	*áwá*	'catch'	*ámá*	'in-law'
/y/	*wììré*	'set out to dry'	*yììré*	'snake'

3.1.4.15 /y/

The palatal glide /y/ becomes nasalized [yⁿ] after a nasal vowel. This would result in its neutralization with intervocalic or word-final singleton /ɲ/, which also surfaces as [yⁿ], though the opportunities for this neutralization are rare. The most common situation in which /y/-nasalization arises is in the formation of deverbal nouns from a monosyllabic verb stem with a nasal vowel, where the second half of the verb's vowel is replaced with [y]. Thus, we see *bóy* 'name' from *bòó* 'call', but *tɔ̌ⁿyⁿ* 'writing' from *tɔ́ɔ́ⁿ* 'write'. The underlying status of the final [yⁿ] as /y/ is clear when we compare it with pairs like *bóy-bòó*. However, certain monosyllabic nouns that are not related to verbs also end in [yⁿ], and in this case it is not clear whether they should be derived from an underlying /VVⁿy/ sequence or a /Vɲ/ sequence, since the [yⁿ] allophone of /ɲ/ would nasalize the preceding vowel.

Take, for example, *těyⁿ* 'worries', where the nasalization marked at the end is realized across the whole syllable. In this case, the more likely underlying form is /těɲ/, with a final palatal nasal, since the underlying form /tèéⁿy/ would contain a [+ATR] nasal vowel /eeⁿ/ that is not part of the vowel inventory (see section 3.2.2). However, it is also possible that this is historically a deverbal noun from a stem *téɛ́ⁿ*, where the noun has undergone a change to [+ATR], an attested change in such noun-verb pairs as *bàá* 'beat (drum)' and *bǒy* 'tom-tom' (see section 13.1.5). In stems like *dɛ́yⁿ* 'different', the situation is even more complicated in that /ɛ́ɛ́ⁿ/ is a licit nasalized vowel, so the space of possibilities for the underlying form is even greater.

Since there is, to my knowledge, no way of distinguishing an underlying /VVⁿy/ sequence from a /Vɲ/ sequence in the absence of morphological information such a noun-verb pair, I will remain neutral on the position, writing such words as *téyⁿ* and *dɛ́yⁿ*, which reflects their surface pronunciation. Note that this confusion can only

arise from words with long vowels (or monosyllabic words with a [yⁿ] coda, which would cause vowel shortening), since nasality is only contrastive on long vowels in Tommo So. That is, words like [kúⁿyⁿɔ́ⁿ] 'squirrel' are unambiguously *kúɲɔ́*, since there are no short nasal vowels in the language that could nasalize /y/.

The following minimal pairs show that /y/ contrasts in nasality and continuancy with /ɲ/ in both initial and medial positions (where /ɲ/ has the allophone [yⁿ]), in backness with /w/, and in continuancy and place with /j/.

(26) Minimal pairs for /y/ vs...

/ɲ/	*yám*	'wasted'	*ɲám*	'difficult'
	kúyɔ́	'first'	[kúyⁿɔ́]	'squirrel'
/w/	*yìì́ré*	'snake'	*wìì́ré*	'set out to dry'
/j/	*yúú*	'millet'	*júú*	'comrade'

3.1.4.16 /r/

Tommo So /r/ (practical orthography) is pronounced as an alveolar tap (IPA [ɾ]). This pronunciation is remarkably consistent; it has no other rhotic allophones. However, due to a process of rhotic dissimilation, /r/ may alternate with [l]. See section 3.7.6 for further discussion.

The following minimal pairs show that continuancy, nasality, and rhoticity are contrastive for /r/. Since liquids cannot appear word-initially, /r/ is only contrastive in non-initial positions.

(27) Minimal pairs for /r/ vs...

/d/	*jàrá*	'knock down'	*jàdá*	'to calculate'
/n/	*gìré*	'watch over'	*gìné*	'house'
/l/	*sárá*	'pay'	*sálá*	'bad'

3.1.4.17 /l/

The phoneme /l/ is a lateral alveolar approximant with only this single allophone. The following minimal pairs show /l/ contrasting for rhoticity, continuancy, and nasality. Like /r/, it cannot appear in word-initial position, and hence all minimal pairs show /l/ word-medially.

(28) Minimal pairs for /l/ vs...

/r/	*sálá*	'bad'	*sárá*	'pay'
/d/	*gùlɔ́*	'vomit'	*gùdɔ́*	'pluck'
/n/	*dɔ̀lɔ́*	'ransom'	*dɔ̀nɔ́*	'sell'

3.2 Vowels

3.2.1 Vowel inventory

Tommo So has seven short vowel phonemes, each of which has an equivalent long vowel. In the practical orthography, long vowels are written as two consecutive vowels (/aa/ for IPA /a:/) since there are no contrasts between long vowels a sequence of two like vowels. Nasalization is also contrastive, though all phonemically nasalized vowels are long; in addition, only the three lowest vowels, /a/ and the [-ATR] mid vowels, have nasalized equivalents. This brings the total number of vowel phonemes up to seventeen, the same number as the consonants.

(29)

Short oral	Long oral	Nasalized
u	uu	
o	oo	
ɔ	ɔɔ	ɔɔ^{n}
a	aa	aa^{n}
ɛ	ɛɛ	ɛɛ^{n}
e	ee	
i	ii	

{e, o} and {ɛ, ɔ} form harmonic sets, to be discussed further in the treatment of vowel harmony in section 3.5.

Long vowels in every stem position are of approximately equal length, averaging around 138ms, a little over double the average length of short vowels, 67ms. These measurements were made based on running speech. In utterance-final position, long vowels are shortened slightly; see section 3.4.4. There is no length distinction in monosyllabic words, due to minimality requirements (section 3.3.2).

3.2.2 Nasal vowels

I asserted above that the seven vowel system that exists in oral vowels collapses into a three vowel system in nasal vowels. We do additionally find a couple cases of high nasalized vowels, but they are restricted to ideophones, like the following:

(30) a. *kìì^{n}-káá^{n}* ‘whooping cough’

b. *búú^{n}-ni* ‘thick (expressive adverbial)’

In addition, the nasalized vowel /ii^{n}/ surfaces in one derived verb *píí^{n}-y^{n}é* ‘get old’. However, this is synonymous with the word *péé^{n}* and could be seen as an irregular derivation in which the diphthong created by vowel hiatus resolution is leveled into a long high vowel (cf. expected *pé-í^{n}y^{n}é*); see section 3.7.3 for more on vowel hiatus with high vowel suffixes.

Given the restricted distribution of these two nasalized vowels, I do not treat them as phonemic. However, it should be noted that all nasalized vowels are extremely rare compared with their oral counterparts. This is especially true for the low vowel /aaⁿ/, whose phonemic status is based on a single regular lexical item *pááⁿ* 'dry up', which forms no true minimal pairs with oral /aa/. The other two nasal vowels, /ɔɔⁿ/ and /ɛɛⁿ/, appear in just five and three non-ideophonic native stems, respectively. These are all listed below:

(31) a. Stems with /ɔɔⁿ/
dɔ́ɔ́ⁿ 'gutter spout'
jɔ̀ɔ́ⁿ 'draw'
pɔ̀ɔ́ⁿ 'fonio'
sɔ́ɔ́ⁿ 'douse (fire)'
tɔ́ɔ́ⁿ 'fill up'

b. Stems with /ɛɛⁿ/
ɛ́ɛ́ⁿ 'tighten; marry; tough'
pɛ́ɛ́ⁿ 'get old'
tɛ́ɛ́ⁿ 'hobble (a donkey)'

As we can see, all nasal vowels are peripheral phonemes in Tommo So, but I take the appearance in at least one regular lexical item as the criterion for determining phonemic status. Given this, we are left with a three-way contrast in nasal vowels, between /aaⁿ/, /ɔɔⁿ/, and /ɛɛⁿ/.

3.2.3 The status of [ə]

In addition to these vowel phonemes, metrically and morphologically conditioned vowel reduction can lead to the creation of schwa [ə] (with a pronunciation closer to high central [ɨ]). For example, in certain trisyllabic verb roots, the medial (second syllable) vowel is pronounced in normal speech with a schwa, as in [ádəbá] 'think'. When asked to pronounce the word slowly, younger speakers offer [ádúbá], but I am told that an older speaker would have said [ádábá]. It appears that what began as a system of vowel reduction has resulted in the restructuring of the UR, with [ə] taken to be an allophone of /i/ or /u/.

This system is challenging in a number of ways. First, not all trisyllabic verbs have undergone this restructuring. *Ádúbá* 'think' has, but consultants give similar verbs like *ábárá* 'spread out' with three /a/ vowels. Second, the system appears to be morphologically constrained in that nouns are much less susceptible to this reduction than verbs. Trisyllabic nominal and adjectival stems like *ságárá* 'young' or *àdàlá* 'half ripe' do not undergo as much second syllable reduction and as a result do not show the same medial vowel restructuring. It is not clear what drives

this asymmetry. Its effects are even felt in deverbal nouns, where a middle-aged consultant gives *ádúbá* for 'think' but both *àdùbú* and *àdàbú* as possible pronunciations of the cognate nominal 'thought (n.)'. Note that in general, [u]-final trisyllabic nouns are more susceptible to vowel reduction than trisyllabic nouns with the same vowel quality in every syllable, as though a final non-high vowel serves to protect a medial one.

This vowel reduction and restructuring probably greatly contributed to large number of /i/-initial suffixes, which delete the final vowel of the stem in vowel hiatus. See section 3.7.3 for further discussion.

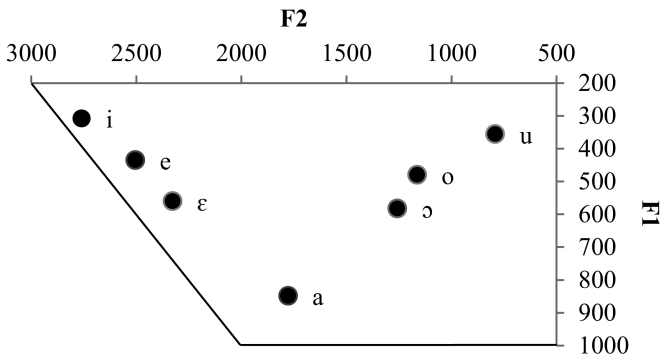
The fact that phonetic [ə] would be reinterpreted as a phonologically high vowel is not surprising, given both the susceptibility of high vowels to consonant effects in the language and the fact that [u] is the default epenthetic vowel. It appears that high vowels are perceived by speakers as being somehow less firm than the non-high vowels, and hence a non-phonemic pronunciation like [ə] can be more easily attributed to high vowels than non-high vowels.

In light of these facts, I treat [ə] as an allophone of /u/ and /i/ and not as a phoneme in and of itself. Nonetheless, much remains to be explained as to why it is so much more pervasive in the verbal system than in the nominal system and why certain verbs have undergone the high vowel reinterpretation while others have been resistant.

3.2.4 Vowel formants

The graph in (32) plots the seven oral vowels with their values of the first and second formant (F1 and F2) in hertz, averaged across the analysis of several words, each repeated three times.

(32) *Tommo So vowel inventory with formant values*[4]



4 u: /tú̲ú, kú̲ú, bú̲dú/. o: /yò̲gó, ò̲béélè, tò̲ndòó, póó, ó̲bó, wó̲/. ɔ: /tɔ́nd̲ɔ̀m, úwɔ=mbe, bɔ̀g̲ɔ́/. a: /à̲gá, yà̲bá, yáá/. ɛ: /ɛ́nɛ́, íg̲ɛ́, dɛ́g̲ɛ́, íb̲ɛ́, ɛ́b̲ɛ́/. e: /òb̲éélè̲, ìsé̲, né̲gu, yàà=bé̲/. i: /íí̲, íb̲ɛ́, ì̲sé/.

We can see that the mid and high vowels are much closer to one another than they are to the low vowel /a/. There may be additional factors helping speakers to distinguish the ATR (advanced tongue root) feature on mid vowels other than simply F1 and F2, but these factors are not audible to my non-native ears.

3.2.5 Minimal pairs

Since Tommo So is a language with strict vowel harmony requirements (section 3.5), minimal pairs for polysyllabic words will almost uniformly have vowel contrasts in multiple positions. For example, we find pairs like *àná* 'rain' and *èné* 'goat', but **àné* or **èná*. A minimal pair is, by definition, two words that differ in only a single segment or feature value, but in Tommo So, a change in the feature value of one vowel is copied onto all other vowels in the word. Since this is only a difference in a single parameter, in this sense *àná* and *èné* are a minimal pair, differing only in vowel quality (/a/ vs. /ɛ/). However, for clarity in the following subsections, I try and present stems with only a single vowel. Monosyllabic C-final stems (with final epenthetic [u], section 3.4.6) are the best location to find such minimal pairs, since they typically only involve one stem vowel. Epenthetic [u] is never used as a place of contrast in the minimal pairs that follow. Minimal pairs contrasting long and short vowels are given only in section 3.2.5.8–3.2.5.14 on long vowels, and minimal pairs contrasting nasal and oral vowels are given only in section 3.2.5.15–3.2.5.17 on nasal vowels.

3.2.5.1 /u/

The vowel /u/ is a high back rounded vowel. In second syllable position in words like *táŋúndá* 'transfer', it has the realization [ə]; see section 3.2.3. The following minimal pairs show that /u/ contrasts in backness (and correlating rounding) with /i/, and in height with /o/ and /ɔ/.

(33)

	Minimal pairs for /u/ vs...			
/i/	*kúdu*	'handle'	*kídu*	'seed residue'
/o/	*kúló*	'hair'	*kóló*	'raw'
/ɔ/	*kúnnɔ́*	'farming huts'	*kɔ́nnɔ́*	'curved'

3.2.5.2 /o/

The vowel /o/ is a mid back rounded vowel. It is additionally specified as being [+ATR]. It contrasts in backness (and rounding) with /e/, in height with /u/ and /a/, and in ATR value with /ɔ/.

(34) Minimal pairs for /o/ vs...

/e/	*sólu*	'cream of millet'	*sélu*	'last bit of liquid'
/u/	*kóló*	'raw'	*kúló*	'hair'
/a/	*tólu*	'pig'	*tálu*	'egg'
/ɔ/	*púló*	'unweave'	*púlɔ́*	'Fulani'

3.2.5.3 /ɔ/

The vowel /ɔ/ is a back mid rounded vowel with a [-ATR] specification. It contrasts with /ɛ/ in backness (and rounding), /u/ and /a/ in height, and /o/ in ATR.

(35) Minimal pairs for /ɔ/ vs...

/ɛ/	*dɔ̀gú*	'quiver'	*dɛ̀gú*	'poverty'
/u/	*kúnnɔ́*	'farming huts'	*kɔ́nnɔ́*	'curved'
/a/	*dɔ̀gú*	'quiver'	*dàgú*	'small'
/o/	*dúgɔ́*	'necklace'	*dúgó*	'poison'

3.2.5.4 /a/

The vowel /a/ is a low central unrounded vowel. It can sometimes be reduced to extra-short [a̯], as in [kádą̯ná] 'oldest man in the village', but the pronunciation of this reduced vowel is lower than that of [ə], the allophone of /u/ and /i/. Given that it is neither front nor back, I contrast it in the following pairs with both front mid /ɛ/ and back mid /ɔ/.

(36) Minimal pairs for /a/ vs...

/ɛ/	*dàgú*	'small'	*dɛ̀gú*	'poverty'
/ɔ/	*dàgú*	'small'	*dɔ̀gú*	'quiver'

3.2.5.5 /ɛ/

The vowel /ɛ/ is a front mid unrounded vowel, specified as [-ATR]. It contrasts with /ɔ/ in backness (and rounding), with /i/ and /a/ in height, and with /e/ in ATR.

(37) Minimal pairs for /ɛ/ vs...

/ɔ/	*dɛ̀gú*	'poverty'	*dɔ̀gú*	'quiver'
/i/	*kɛ́dɛ́*	'cut'	*kídɛ́*	'thing'
/a/	*dɛ̀gú*	'poverty'	*dàgú*	'small'
/e/	*bɛ̀lú*	'animal'	*bèlú*	'grass'

3.2.5.6 /e/

The vowel /e/ is a front mid unrounded vowel with a [+ATR] specification. It contrasts with /o/ in backness (and rounding), with /i/ and /a/ in height, and with /ɛ/ in ATR.

(38) Minimal pairs for /e/ vs...

/o/	*sélu*	'last bit of water'	*sólu*	'cream of millet'
/i/	*pédé*	'shell'	*pídé*	'spray'
/a/	*bàré*	'meeting'	*bàrá*	'help'
/ɛ/	*bèlú*	'grass'	*bɛ̀lú*	'animal'

3.2.5.7 /i/

The vowel /i/ is a high front unrounded vowel. It has the allophone [ə] in the metrically weak second syllable position in verbs like *tɛ́ŋíndɛ́* [tɛ́ŋə́ndɛ́] 'turn on lights'. In addition, slightly lower /i/ ([i̟]) has been found even in first syllable position in words such as *bìl-áa* 'became'. This could be due to the following low vowel, since words like *bílu* 'ladder' do not show this same lowering effect. More data are required to properly sort out the phonetic realization of /i/. It contrasts in backness (and rounding) with /u/ and in height with /e/ and /ɛ/.

(39) Minimal pairs for /i/ vs...

/u/	*kídu*	'seed residue'	*kúdu*	'handle'
/e/	*pídé*	'spray'	*pédé*	'shell'
/ɛ/	*kídɛ́*	'thing'	*kɛ́dɛ́*	'cut'

3.2.5.8 /uu/

Unlike short /u/, long /uu/ never reduces to schwa. All long vowels have a half-long allophone in utterance final position, rare in naturalistic speech; see section 3.4.4. /uu/ is contrastive with /ii/ in backness (and rounding), with /oo/ and /ɔɔ/ in height, and with /u/ in length.

(40) Minimal pairs for /uu/ vs...

/ii/	*dúú*	'below'	*díí*	'water'
/oo/	*yúú*	'millet'	*yóó*	'enter'
/ɔɔ/	*túúrɔ́*	'spit lightly'	*tɔ́ɔ́rɔ́*	'let do'
/u/	*kúúló*	'harvest corn'	*kúló*	'hair'

3.2.5.9 /oo/

The long vowel /oo/ is contrastive with /ee/ in backness (and rounding), with /uu/ and /aa/ in height, with /ɔɔ/ in ATR, and with /o/ in length.

(41) Minimal pairs for /oo/ vs...

/ee/	*dòó*	'millet type'	*dèé*	'know'
/uu/	*yóó*	'enter'	*yúú*	'millet'
/aa/	*bòó*	'call'	*bàá*	'beat (drum)'
/ɔɔ/	*dòó*	'millet type'	*dɔ̀ɔ́*	'arrive'
/o/	*wòòló*	'kola nut'	*wòló*	'cave in'

3.2.5.10 /ɔɔ/

The long vowel /ɔɔ/ is contrastive with /ɛɛ/ in backness (and rounding), with /uu/ and /aa/ in height, with /oo/ in ATR, and with /ɔ/ in length.

(42) Minimal pairs or /ɔɔ/ vs...

/ɛɛ/	*sɔ́ɔ́*	'speak'	*sɛ́ɛ́*	'beer residue'
/uu/	*tɔ́ɔ́rɔ́*	'let do'	*túúrɔ́*	'spit lightly'
/aa/	*dɔ̀ɔ́*	'arrive'	*dàá*	'kill'
/oo/	*sɔ̀ɔ́*	'speech'	*sòó*	'sweat'
/ɔ/	*bɔ́ɔ́ndu*	'pampered'	*bɔ́ndu*	'marrow'

3.2.5.11 /aa/

The long vowel /aa/ contrasts in height with both /ɔɔ/ and /ɛɛ/, and in length with /a/.

(43) Minimal pairs for /aa/ vs...

/ɔɔ/	*sáá*	'wild grape'	*sɔ́ɔ́*	'to speak'
/ɛɛ/	*sáá*	'wild grape'	*sɛ́ɛ́*	'beer residue'
/a/	*sàdáá*	'bird'	*sàdá*	'garden'

3.2.5.12 /ɛɛ/

The long vowel /ɛɛ/ contrasts with /ɔɔ/ in backness (and rounding), with /ii/ and /aa/ in height, with /ee/ in ATR, and with /ɛ/ in length.

(44) Minimal pairs for /ɛɛ/ vs...

/ɔɔ/	*sɛ́ɛ́*	'beer residue'	*sɔ́ɔ́*	'to speak'
/ii/	*pɛ́ɛ́ndɛ́*	'line up'	*pííndé*	'close'
/aa/	*sɛ́ɛ́*	'beer residue'	*sáá*	'wild grape'
/ee/	*nɛ́ɛ́*	'now'	*néé*	'two'
/ɛ/	*pɛ́ɛ́ndɛ́*	'line up'	*pɛ́ndɛ́*	'sores'

3.2.5.13 /ee/

The long vowel /ee/ contrasts with /oo/ in backness (and rounding), /ii/ and /aa/ in height, /ɛɛ/ in ATR, and /e/ in length.

(45) Minimal pairs for /ee/ vs...

/oo/	*dèé*	'know'	*dòó*	'millet type'
/ii/	*déé*	'Papa!'	*díí*	'water'
/aa/	*néé*	'two'	*náá*	'mother'
/ɛɛ/	*néé*	'two'	*nɛ́ɛ́*	'now'
/e/	*tééndé*	'align'	*téndé*	'witness'

3.2.5.14 /ii/

The long vowel /ii/ is not subject to allophonic variation, unlike its short counterpart. It contrasts with /uu/ in backness (and rounding), with /ee/ in height, and with /i/ in length.

(46) Minimal pairs for /ii/ vs...

/uu/	*díí*	'water'	*dúú*	'below'
/ee/	*díí*	'water'	*déé*	'Papa!'
/i/	*kíílé*	'pull off'	*kílé*	'extra hard'

3.2.5.15 /ɔɔⁿ/

The nasal vowel /ɔɔⁿ/ contrasts with /ɛɛⁿ/ in backness (and rounding), /aaⁿ/ in height, and /ɔɔ/ in nasality.

(47) Minimal pairs for /ɔɔⁿ/ vs...

/ɛɛⁿ/	*pɔ́ɔ́ⁿ*	'bridge'	*pɛ́ɛ́ⁿ*	'get old'
/aaⁿ/	*pɔ́ɔ́ⁿ*	'bridge'	*pááⁿ*	'dry up'
/ɔɔ/	*sɔ́ɔ́ⁿ*	'extinguish'	*sɔ́ɔ́*	'speak'

3.2.5.16 /aaⁿ/

The nasal vowel /aaⁿ/ is vanishingly rare, as noted above in section 3.2.2. The only regular (i.e. native and non-ideophonic) lexical item containing it is *pááⁿ* 'dry up'. The following minimal or near-minimal pairs show that /aaⁿ/ contrasts with /ɛɛⁿ/ and /ɔɔⁿ/ in height and with /aa/ in nasality.

(48) Minimal pairs for /aaⁿ/ vs...

/ɛɛⁿ/	*pááⁿ*	'dry up'	*pɛ́ɛ́ⁿ*	'get old'
/ɔɔⁿ/	*pááⁿ*	'dry up'	*pɔ́ɔ́ⁿ*	'bridge'
/aa/	*pááⁿ*	'dry up'	*báá*	'father'

3.2.5.17 /ɛɛⁿ/

The nasal vowel /ɛɛⁿ/ contrasts in backness (and rounding) with /ɔɔⁿ/, in height with /aaⁿ/, and in nasality with /ɛɛ/.

(49) Minimal pairs for /ɛɛⁿ/ vs...

/ɔɔⁿ/	*pɛ́ɛ́ⁿ*	'get old'	*pɔ́ɔ́ⁿ*	'bridge'
/aaⁿ/	*pɛ́ɛ́ⁿ*	'get old'	*pááⁿ*	'dry up'
/ɛɛ/	*pɛ́ɛ́ⁿ*	'get old'	*pɛ́ɛ́*	'squash'

3.3 Syllable and stem structure

3.3.1 Syllable shape

Tommo So syllables can be one of the following shapes: (C)V, (C)V:, (C)VR (where R stands for a sonorant), (C)VC (where the coda is the first half of a geminate), N, NCV, NCV: or CVV (with vowel hiatus). The last type arises almost exclusively in verbal inflection, but is present in some loanwords as well.

The following provides examples of each type of syllable:

(50)

a.	V	*è.né*	'goat'
b.	CV	*gì.né*	'house'
c.	V:	*íí*	'child'
d.	CV:	*nàá*	'cow'
e.	VR	*ém*	'milk'
f.	CVR	*năm*	'sun'
g.	CVC	*sɔ́b.bɔ̀*	'dry sowing'
h.	N	*ɲ́.yɛ́*	'eat'
i.	NCV	*à.ndá*	'udder'
j.	NCV:	*gà.mbáá*	'some'
k.	CVV	*dɔ̀è*	'he arrived'

Of these, CVC and N are the rarest, since non-sonorant geminates are rare, present only in loanwords, and syllabic nasals are found at the beginning of only a handful of words (syllabic /m/ 3 stems, syllabic /n/ 5 stems, syllabic /ɲ/ 4 stems). CVR syllables occur almost exclusively in monosyllabic stems, but these often surface as disyllabic with an epenthetic [u] following the sonorant (yielding CV.CV); see section 3.4.6 for more on epenthesis.

CVV syllables in which the two V slots are not of the same quality are not legal in native stems, though they can be created by inflection. They most commonly arise in the defocalized perfective, which is marked with a suffix *-e ~ -ɛ* or *-i*. Vowel hiatus following a short vowel is resolved by deleting the stem-final vowel (*dɔ̀nɔ̀-è* → *dɔ̀n-è*

'sold'), but with a long vowel, only the second half of the vowel is deleted (*dɔ̀ɔ̀-ɛ̀* → *dɔ̀-ɛ̀* 'arrived'), meaning that vowel hiatus remains on the surface. These two vowels are pronounced as one syllable, with the two vowels generally forming a diphthong. For instance, in *dɔ̀-ɛ̀* 'he arrived', /ɔ/ shortens to almost a mid back glide [dɔ̯̀ɛ̀]. Vowel hiatus can also arise from the addition of high-vowel suffixes; see section 3.7.3.

Vowel hiatus across syllable boundaries (V.V) is very rare in Tommo So and does not occur underlyingly. However, medial V syllables can be created when the transitive suffix is added to a verb whose final syllable is /w/-initial. The change proceeds as follows:

(51) /gàwá + irɛ/ → |gàwírɔ́| → [gàúrɔ́] 'entrust (sth) to (sb)'

We know that the root is underlyingly /gàwá/ due to the existence of a mediopassive verb sharing the same stem *gàwá-íyɛ́* 'guard (sth) for (sb)'. The [wi] syllable created by vowel hiatus with a high-vowel suffix is realized on the surface as [u]. Other verbs share this surface [a.u] pattern, with a medial [u] syllable, but without the support of another verb pointing to its underlying form. Nonetheless, they are all of the form [(C)aurɔ], and so it seems likely that, underlyingly, they are derived from this same sequence. Other than these derived [u], I know of no medial V syllables.

There is some question as to the treatment of NC clusters in words such as *jàŋgá* 'study'. In word-initial position, we find only syllabic nasals followed by a stop, rather than prenasalized stops. This is evidenced by the fact that the word-initial nasal (when followed by a consonant, not when by itself as an onset) is always a tone-bearing unit, indicating that it is moraic, unlike the nasal of a prenasalized stop. For instance:

(52) *ɲ́.yɛ̀-dɛ̀-m*
eat-IMPF-1SG
'I eat'

In the imperfective aspect, the verb stem has a {HL} tone melody, with H on the first mora only. In this case, the H is assigned to /ɲ/, showing its moraic nature. Word-final nasals also add a mora, as evidenced by the fact that CVN syllables are able to carry a contour tone, as in *nă̌m* 'sun', *yĭ̌m* 'death', etc.

It is less clear whether *jàŋgá* 'study' should be syllabified as /jàŋ.gá/ with a coda nasal or as /jà.ŋgá/ with a prenasalized stop. Evidence differentiating the two analyses is hard to come by. First, we could ask whether the CVN syllable in a CVNCV word is able to carry a contour tone, which would indicate its bimoraic nature. In fact, we do not find any contour tones on these syllables, but Tommo So disallows rising tones on non-final syllables (see section 4.1.1), and falling tones are not present in native stems; thus, the lack of contour tones here may be independent of syllable structure. We could also look to grammatical tone overlays that are

sensitive to mora count. For example, in inalienable possession with a pronominal possessor, the possessed noun will receive a {H} overlay if it is two moras or less and a {HL} overlay otherwise. We could look, then, to see how a CVNCV word treated in this possessive paradigm. The only inalienable noun of this shape is *níɲju* 'maternal uncle', which receives a {H} overlay, suggesting a bimoraic weight. Once again, though, this example is confounded by the fact that the final [u] is epenthetic and hence not part of the underlying structure. Interestingly, then, this means that the grammatical tone overlay is assigned based on the *underlying* number of moras, not the surface number, assuming that an epenthetic vowel would contribute a mora. For more on grammatical tone, see section 4.4–4.5. The one piece of evidence that points towards a prenasalized stop analysis (syllable shape NCV, the analysis pursued here) is that we find long vowels occurring before a nasal stop cluster in a word, as in *dùù-ndó* 'put down' or *nííndɛ́* 'tongue'. Since Tommo So disallows super-heavy syllables (syllables with more than two moras), we can deduce that the nasal in this case does not add a mora. The asymmetry between initial nasal stop clusters and medial clusters, wherein the initial nasals are tone bearing and syllabic while the medial ones are not, can be explained by a constraint against falling sonority word onsets: a nasal is more sonorous than a stop, so the sonority decreases as one moves into the syllable. This is tolerated word-medially but not in word-initial position, and instead those clusters are broken up into two syllables.

3.3.2 Word minimality requirement

Like many languages, Tommo So imposes a bimoraic word minimum. Thus, we find words like *díí* 'water' and *kúú* 'head', but no monomoraic equivalent (**dí*, **kú*). The bimoraic size allows all stems to carry the language's lexical tonal melodies, /H/ and /LH/; a monomoraic stem could not host the /LH/ melody.

Nearly all stems meet the minimality requirement, except for three exceptional verb stems: *gɛ̀* 'say', *yɛ̀* 'see', and *jɛ̀* 'take'. In most inflections, suffixes can be added to bring the stem up to word size, but in forms like the defocalized perfective, we may see subminimal words like *g-ì* '(s)he said' from *gɛ̀* 'say'.

Note that we can use this measure of minimality to distinguish independent words from clitics, which are generally monomoraic. See section 3.8 for further discussion.

3.3.3 Stem lengths

All known stems are between one and three syllables in length, with monosyllabic stems being almost exceptionlessly bimoraic (either a long vowel or a sonorant coda for surface monosyllabic stems); the exceptions are listed in section 3.3.2 above. Trisyllabic stems are rarer than their mono- and disyllabic counterparts. Below are examples of stems of different lengths:

(53)	a.	Monosyllabic	(1 mora)	*gɛ̀*	‘say’
				yɛ̀	‘see’
	b.	Monosyllabic	(2 moras)	*nằm*	‘sun’
				sɔ́ɔ́	‘speak’
	c.	Disyllabic	(2 moras)	*ìsé*	‘dog’
				dàmá	‘taboo’
	d.	Disyllabic	(3 moras)	*dámmá*	‘village’
				gèèdé	‘thin cotton thread’
	e.	Disyllabic	(4 moras)	*gààlúú*	‘last year’
				tànnàá	‘cane’
	f.	Trisyllabic	(3 moras)	*kágádá*	‘sear’
				bùgùdɔ́	‘chubby’
	g.	Trisyllabic	(4 moras)	*dénnélé*	‘circle’
				mòòmíyó	‘scorpion’

Section 3.3.1 above brought up the case of underlyingly C-final stems that surface with a final epenthetic [u], in that case *nŋ̍ju* ‘maternal uncle’. It seems that cases such as these should be considered an underlying monosyllabic stem with two moras, based on evidence from the tonal system, even though it surfaces as disyllabic with the addition of the epenthetic vowel. For further discussion of vowel epenthesis, see section 3.4.6.

3.3.4 “Crypto-compounds”

Most, if not all, long words (four syllables or more) behave phonologically as though they were compounds rather than a single stem, even if they contain no recognizable smaller stems. Following Heath (2008), I call such words **crypto-compounds**, a compound made up of lexically unidentifiable parts. These compounds are like English *cranberry*, where *cran-* is a bound root, only in the case of Tommo So crypto-compounds, both alleged roots are bound roots.

Crypto-compounds often have a prosodically-marked juncture, generally but not always split into two trochees [ó.ŏ-ó.ŏ]. They typically exhibit one of two tonal patterns characteristic of compounds: either the first “root” receives a {L} overlay replacing its lexical tone or the second “root” does. These parallel canonical compounds and pseudo-genitive compounds, respectively, treated in Chapter 6. Nonetheless, some common crypto-compounds like /bándáŋ-kálá/ ‘courtyard’ show that they can be all /H/ as well.

Apart from prosodic cues and overall length, there are segmental cues that these words should be treated as compounds, opaque though they may be. Turning again to ‘courtyard’, the crypto-compound status of the word is made clear by the word-internal sequence /ŋk/. The phonotactics prohibits voiceless plosives stem-internally

(see section 3.4.1.2), so this series of consonants must be licensed by some sort of prosodic juncture.

Below is a list of features that may distinguish crypto-compounds from single stems:

(54) a. *Tone pattern* of either [σσ-ȍȍ] or [ȍȍ-σσ], the unmarked tones generally being /H/ or /LH/, the two lexical tonal melodies.
Ex. *èŋèlè-gèèndɛ́* 'aloe'
dúrúm-bàànà 'amber'

b. *Separate harmonic domains*, where vowels in the first half of the word do not agree in backness or ATR with the vowels in the second half.
Ex. *èndɛ̀-kúmɔ́* 'centipede'

c. *Word-internal voiceless plosive*, or other phonotactic violations.
Ex. bándáŋ-kálá 'courtyard'

The vast majority of crypto-compounds are found in the names of plants and animals, and it seems likely that the crypto-compounds were at one point transparent. In fact, we find certain bound roots repeating in multiple crypto-compounds, though they lack any synchronic meaning of their own. For example, the bound root *èndɛ̀* seen in 'centipede' in (54b) above is also found in human names, like *èndɛ̀-kìndíyé* and in other insect names like *èndɛ̀-kɛ̌w* 'ant'. This suggests that it may have at one point carried its own meaning.

In addition to full crypto-compounds, composed of two bound roots, half opaque compounds also exist in the lexicon. These parallel English compounds like *cranberry* in that half of the compound (in the English case, *berry*) is a free root with identifiable meaning, but it combines with another root unable to stand on its own. The same phonological cues listed in (54) also apply to half-opaque compounds, which tend to be more common and are frequently found in everyday vocabulary. The following gives examples with bound roots (glossed as '?') in both initial and final position in the compound:

(55) a. *sànà*^{L} *èmmɛ́*
? sorghum
'maize'

b. *ànù*^{L} *tɔ̀dú*
leg ?
'calf'

Sometimes the same bound root can be used in many different compounds, which can help to pinpoint the original meaning of the bound root. This is the case with *tàbà*, found in compounds like *tàbà*^{L} *dánná* 'large, flat boulder' (cf. *dánná* 'outside')

or *tàbà*^{L} *kámbé* ‘cooking coarsely crushed grain between hot stones’ (cf. *kámbá* ‘cook coarsely crushed grain between hot stones’), suggesting that the original root *taba* might have had a meaning related to ‘stone’. Nonetheless, other compounds like *tàbà*^{L} *dìì*^{L} *kùndú* ‘gourd for carrying water’ (cf. *díí kúndó* ‘put water’) obscure this sense.

3.4 Phonotactics

Sections 3.1 and 3.2 introduced the reader to the individual consonants and vowels of Tommo So. In this section, I discuss phonotactic restrictions on where in the stem phonemes can occur and how they can combine with one another.

3.4.1 Constraints on individual consonants

3.4.1.1 Word-initial restrictions

In word-initial position, the liquids /r/ and /l/ are not allowed. This restriction is stronger for /r/ in that loanwords with initial /r/ tend to be repaired by epenthesis, which is not true for /l/. Unlike in final position, however, the epenthetic vowel in initial position is a copy of the following vowel, as in:

(56) a. *Ramata* (female name) → *aràmátá*
b. *réunion* ‘meeting’ → *ɛrénúyón*

For more on vowel epenthesis, see section 3.4.6. Only one /r/-initial word is left unrepaired, and that is the ideophone *rék*, an intensifier for the number ‘one’.

The liquid /l/ never occurs in word-initial position in native stems, but its presence in the onset of loanwords is usually left unrepaired. Thus, for French *lettre* ‘letter’, we see Tommo So *léètèrè*. However, if a loanword is sufficiently native-looking, it may be further assimilated into the native vocabulary by replacing /l/ with /n/. Consider first *léètèrè* ‘letter’. This word is non-native in a number of ways. First, it begins with a liquid. Second, it contains a voiceless stop /t/ word-internally. Third, it has a {HL} tone pattern instead of either {H} or {LH}, the only native tonal melodies on lexical stems. All of these phonological cues tell the speaker that this word is not part of the native lexicon. In a word like *làmbá* ‘lamp’ (from French *lampe*), on the other hand, everything but the onset looks native, and it forms a minimal pair with native words like *dàmbá* ‘push’. In cases like these, it seems that some speakers bring the stem into the native lexicon, which then forces a repair of the liquid onset, yielding a pronunciation *nàmbá*. Note that Tommo So enclitics may begin with /l/; these are not barred because the restriction is on word-initial position.