SUPPLEMENTARY MATERIALS FOR
BHASA
INDONESIAN MINIMAL PAIRS

This document serves as a supplement to the paper "BHASA: A Holistic Southeast Asian Linguistic and Cultural Evaluation Suite for Large Language Models" and the handcrafted data that comes with it. It contains a single pair of examples (grammatical and ungrammatical) for each category (and important subcategories) of minimal pairs in the Indonesian LINDSEA dataset.

Each grammatical sentence is glossed with 4 or 5 lines: (1) The original sentence, (2) the sentence with morphemic boundaries shown, (3) linguistic glossing, (4) the literal meaning of the sentence (where relevant), and (5) the intended meaning of the sentence. We only provide the original sentence for ungrammatical sentences since it is sometimes not feasible to provide a meaning for a sentence that is not well-formed, and morphemic boundaries can be inferred from the grammatical sentence given the minimal differences between them. Note that the term 'ungrammatical' is used to indicate that the sentence is incorrect in the context of the linguistic phenomenon being studied (since there may be some cases where the sentence is not, strictly speaking, ungrammatical). Comments have been provided where necessary to further clarify the intent behind the category/subcategory.

Linguistic Phenomenon	Category	Subcategory	Grammatical Sentence	Ungrammatical Sentence	Comments
Morphology	Men- with clefted sentences	-	Jendela itulah yang Tono rusak / Jendela itu-lah yang Tono rusak / window DEM-LAH REL Tono destroy / 'That window is the one that Tono destroyed.'	Jendela itulah yang Tono merusak	Men- is ungrammatical on verb when object is clefted
	Men- with control verbs mau and coba	Coba	Rumah itu Tono coba beli / Rumah itu Tono coba beli / house DEM Tono try buy / Lit.: "That house, Tono tried to buy." / 'Tono tried to buy that house.'	Rumah itu Tono coba membeli	Men- is ungrammatical on main verb in object voice construction
		Mau	Uang itu Tono mau tabung / Uang itu Tono mau tabung / money DEM Tono want save / Lit.: "That money, Tono wants to save." / 'Tono wants to save that money.'	Uang itu Tono mau menabung	Men- is ungrammatical on main verb in object voice construction
	Men- with dirinya/dirimu subjects	-	Dirimu kamu ubah / Dirimu kamu ubah / SELF.2SG 2SG change / 'You yourself change.'	Dirimu kamu mengubah	Men- is ungrammatical with dirinya/dirimu subjects
	Men- with objects subject to Left Branch Condition	WH-in-situ	Tono mencari mobilnya siapa? / Tono men-cari mobil-nya siapa? / Tono MEN-search car=POSS who / Lit.:"Tono searches for whose car?" / 'Whose car did Tono search for?'	Tono cari mobilnya siapa?	Only grammatical with Men- on verb
		WH-movement from island	Siapa yang Tono baca bukunya? / Siapa yang Tono baca buku=nya? / who COMP Tono read book=POSS / Lit.: "Whose did Tono read book of?" / 'Whose book did Tono read?'	Siapa yang Tono membaca bukunya?	Ungrammatical with Men- on verb
	Adjective Reduplication	-	Salah satu dari tas-tas kecil itu punya saya? / Salah.satu dari tas-tas kecil itu punya saya? / one.of from bag-RED small DEM HAVE 1SG / 'One of those small bags is mine.'	Salah satu dari tas-tas kecil-kecil itu punya saya	Adjective cannot be reduplicated in this context
	Noun Reduplication (Endocentric Compounds)	-	Buku-buku sejarah itu punya saya / Buku-buku sejarah itu punya saya / book-RED history DEM HAVE 1SG / 'Those history books are mine.'	Buku sejarah-sejarah itu punya saya	Only the head of the compound can be reduplicated
	Noun Reduplication (Plurals)	-	Komputer-komputer canggih itu punya saya / Komputer-komputer canggih itu punya saya / computer-RED advanced DEM HAVE 1SG / 'Those advanced computers are mine.'	Komputer canggih-canggih itu punya saya	Only the noun can be reduplicated
	Benefactive constructions without untuk	-	Tono menuliskan Tini puisi cinta / Tono men-tulis-kan Tini puisi cinta / Tono MEN-write-KAN Tini poem love / 'Tono wrote Tini love poems.'	Tono menuliskan puisi cinta Tini	Untuk can be left out but word order is no longer as flexible
	Dative constructions without kepada	-	Tono memberikan Tini uang / Tono men-beri-kan Tini uang / Tono MEN-give-KAN Tini money / 'Tono gave Tini money.'	Tono memberikan uang Tini	Kepada can be left out but word order is no longer as flexible
	Preposition stranding	-	Di mana Tono pergi makan malam hari ini? / Di mana Tono pergi makan malam hari ini? / at where Tono go eat night day DEM / 'Where did Tono go to eat dinner today?'	Mana Tono pergi makan malam hari ini di?	Prepositional stranding is ungrammatical in Indonesian
	Correlative conjunctions	-	Walaupun Tono malas, tapi dia pintar / Walaupun Tono malas, tapi dia pintar / even.though Tono lazy, but 3SG smart / 'Even though Tono is lazy, he is smart.'	Walaupun Tono malas, dia pintar	In formal Indonesian, the correlative conjunction must be present in both clauses
	Complementizers (bahwa/kalau)	Ungrammatical with bahwa	Siapa yang Tono katakan membeli sebuah mobil baru kemarin? / Siapa yang Tono kata-kan men-beli sebuah mobil baru kemarin? / who COMP Tono say-KAN MEN-buy CL car new yesterday / 'Who did Tono say bought a new car yesterday?'	Siapa yang Tono katakan bahwa membeli sebuah mobil baru kemarin?	Complementizer bahwa cannot be inserted
		Ungrammatical with kalau	Siapa yang Tono sangkal membeli sebuah mobil baru kemarin? / Siapa yang Tono sangkal men-beli sebuah mobil baru kemarin? / who COMP Tono deny MEN-buy CL car new yesterday / 'Who did Tono deny bought a new car yesterday?'	Siapa yang Tono sangkal kalau membeli sebuah mobil baru kemarin?	Complementizer kalau cannot be inserted

Category	Phenomenon	Variant	Interlinear gloss	Paired sentence	Rule
	-il-kan suffix	Grammatical with -i	Tono akhirnya dapat mengenali / Tono akhirnya dapat men-kenal-i / Tono finally can MEN-recognize-I culprit PEN-steal-AN DEM / 'Tono can finally recognize the culprit of the theft.'	Tono akhirnya dapat mengenalkan pelaku pencurian itu	
		Grammatical with -kan	Tono berupaya untuk mengobatkan Tini pada dokter itu / Tono ber-upaya untuk men-obat-kan Tini pada dokter itu / Tono BER-effort for MEN-medicine-KAN Tini to doctor DEM / Lit.: 'Tono made an effort to cure Tini to that doctor.' 'Tono made an effort to cure Tini by sending her to the doctor.'	Tono berupaya untuk mengobati Tini pada dokter itu	
	Verb + -kah	-	Belum datangkah Tono sejak kemarin sore? / Belum datang-kah Tono sejak kemarin sore? / not.yet come-Q Tono since yesterday afternoon / Lit.: 'Hasn't Tono come yet since yesterday afternoon?' 'Hasn't Tono been here since yesterday afternoon?'	Belum datang Tono sejak kemarin sore?	Question particle -kah must be affixed to verb in yes/no questions
Argument Structure	Dirimu sendiri with dia + Dirinya sendiri with kamu	-	Dirimu sendiri harus kamu ajukan sebagai calon ketua kelas / Dirimu sendiri harus kamu ajukan sebagai calon ketua kelas / SELF.2SG SELF must 2SG submit as candidate head class / 'You must volunteer yourself as class prefect.'	Dirimu sendiri harus dia ajukan sebagai calon ketua kelas	Person feature mismatched in ungrammatical sentence
	Dirimu sendiri with/without kamu + Dirinya sendiri with/without dia	-	Dirimu sendiri perlu kamu perbaiki / Dirimu sendiri perlu kamu per-baik-i / SELF.2SG SELF need 2SG PER-good-I / Lit.: 'You need to fix yourself.' 'You need to get it together.'	Dirimu sendiri perlu perbaiki	Pronoun with matching person feature must be present
	Dirimu with dia + Dirinya with kamu	-	Dirimu belum kamu ajarkan untuk berjuang / Dirimu belum kamu ajar-kan untuk ber-juang / SELF.2SG not.yet 2SG teach-KAN for BER-struggle / 'You have not yet learned to fight for yourself.'	Dirimu belum dia ajarkan untuk berjuang	Person feature mismatched in ungrammatical sentence
	Dirimu with/without kamu + Dirinya with/without dia	-	Dirimu mesti kamu serahkan ke polisi / Dirimu mesti kamu serah-kan ke polisi / SELF.2SG must 2SG surrender-KAN to police / 'You must surrender yourself to the police.'	Dirimu mesti serahkan ke polisi	Pronoun with matching person feature must be present
	Modals in Active Voice	-	Tono akan men-beli buku di toko buku / Tono will MEN-buy book to store book / 'Tono will buy books from the bookstore.'	Akan Tono membeli buku di toko buku	Modal must immediately precede verb
	Modals in Object Voice	With -kan on verb	Kue bolunya masih Tono panggangkan buat Tini / Kue.bolu=nya masih Tono panggang-kan buat Tini / sponge.cake=NYA still Tono bake-KAN for Tini / Lit.: 'Those sponge cakes, Tono still bakes for Tini.' 'Tono still bakes those sponge cakes for Tini.'	Kue bolunya Tono masih panggangkan buat Tini	Modal cannot be between verb and subject
		Without -kan on verb	Tasnya masih Tono bawa ke sekolah / Tas=nya masih Tono bawa ke sekolah / bag=NYA still Tono carry to school / Lit.: 'That bag, Tono still carries to school.' 'Tono still carries that bag to school.'	Tasnya Tono masih bawa ke sekolah	Modal cannot be between verb and subject
	Temporal nouns + -lah	-	Dua hari lagilah Tono akan menyewa bus untuk pergi ke Semarang / Dua hari lagi-lah Tono akan men-sewa bus untuk pergi ke Semarang / two day more-LAH Tono will MEN-rent bus for go to Semarang / 'Two days later, Tono will rent a bus to go to Semarang.'	Tono akan menyewa bus untuk pergi ke Semarang dua hari lagilah	Constituent suffixed with -lah must be fronted
	Di-VP topicalization with =nya	-	Ditamparnya Tono dengan tangannya / Di-tampar=nya Tono dengan tangan=nya / DI-slap=NYA Tono with hand=NYA / Lit.: 'Tono was slapped with his/her hand.' 'Tono was slapped by someone's hand.'	Ditampar Tono dengan tangannya	
	Ada	With =nya	Ada lutung di dapur Tono / Ada lutung di dapur Tono / COP leaf.monkey in kitchen Tono / 'There is a leaf monkey in Tono's kitchen.'	Ada lutungnya di dapur Tono	Cannot have definiteness marked by =nya with ada preceding subject
		With itu	Lawar itu ada di rumah Tono / Lawar itu ada di rumah Tono / lawar DEM COP in house Tono / 'The lawar is in Tono's house.'	Lawar ada di rumah Tono	Must have definiteness marked by itu with subject preceding ada

Phenomenon	Condition	Example	Simplified	Notes
Asymmetry between animates and inanimates in passive voice	-	Rumahnya kehujanan batu dari bukit / Rumah=nya ke-hujan-an batu dari bukit / house=NYA KE-rain-AN stone from hill / Lit.: 'The house was rained on with stones from the hill.' / 'A landslide from the hill descended upon the house.'	Rumahnya kehujanan oleh batu dari bukit	*Oleh* is ungrammatical in passive constructions with inanimate agents
No *oleh* in passive voice	-	Tono dibelikan Tini kue lapis Surabaya / Tono di-beli-kan Tini kue lapis Surabaya / Tono DI-buy-KAN Tini cake layer Surabaya / Lit.: 'Tono was bought (by) Tini a Surabaya layer cake.' / 'Tono was bought a Surabaya layer cake by Tini.'	Tono dibelikan kue lapis Surabaya Tini	*Oleh* can be left out but word order is no longer as flexible
Oleh-stranding in passive voice	No additional argument	Oleh Tono, Tini dimarahi / Oleh Tono, Tini di-marah-i / by Tono Tini DI-angry-I / Lit.: 'By Tono, Tini was angered.' / 'Tini was angered by Tono.'	Tono, Tini dimarahi oleh	*Oleh* cannot be stranded
	With additional argument	Oleh Tono, Tini dibuatkan kopi / Oleh Tono, Tini di-buat-kan kopi / by Tono Tini DI-make-KAN kopi / Lit.: 'By Tono, Tini a coffee.' / 'Tini was made a coffee by Tono.'	Tono, Tini dibuatkan kopi oleh	*Oleh* cannot be stranded
Demonstratives on objects in object voice	-	Promosi itu Tono perlu dapatkan / Promosi itu Tono perlu dapat-kan / promotion DEM Tono need get-KAN / Lit.: 'That promotion, Tono needs to get.' / 'Tono needs to get that promotion.'	Promosi Tono perlu dapatkan	Must have definiteness marked by *itu*
Possessive =*nya*	-	Tono memakan rendang daging sapinya Tini di dapur / Tono men-makan rendang daging sapi=nya Tini di dapur / Tono MEN-eat rendang meat cow=NYA Tini in kitchen / 'Tono eats Tini's beef rendang in the kitchen.'	Tono memakan rendang daging sapi Tini di dapur	
Word order in benefactive constructions	-	Tono memasakkan rendang untuk Tini / Tono men-masak-kan rendang untuk Tini / Tono MEN-cook-KAN rendang for Tini / 'Tono cooked rendang for Tini.'	Tono memasakkan untuk Tini rendang	
Word order in object voice	With -*kan* on verb	Tehnya buat Tini Tono buatkan / Teh=nya Tono buat-kan buat Tini / tea=NYA Tono do-KAN for Tini / Lit.: 'The tea, Tono made for Tini.' / 'Tono made the tea for Tini.'	Tehnya buat Tini Tono buatkan	
	Without -*kan* on verb	Botolnya Tono isi di dapur Tini / Botol=nya Tono isi di dapur Tini / bottle=NYA Tono fill in kitchen Tini / Lit.: 'The bottle, Tono filled in Tini's kitchen.' / 'Tono filled the bottle in Tini's kitchen.'	Botolnya di dapur Tini Tono isi	
Verb phrase ellipsis	Aspect mismatch	Tono sudah pergi, tapi Tini belum / Tono sudah pergi, tapi Tini belum / Tono already go, but Tini not.yet / Lit.: 'Tono has already gone, but Tini not yet.' / 'Tono has already left, but Tini has not left yet.'	Tono sudah pergi, tapi Tini masih	
	Word order	Tono sedang tidur dan Tini juga masih / Tono sedang tidur dan Tini juga masih / Tono CONT sleep and Tini also still / Lit.: 'Tono is sleeping and Tini also still.' / 'Tono is sleeping and Tini is also still sleeping.'	Tono masih tidur dan Tini juga sedang	
WH-movement with Left Branch Condition	-	Siapa yang Tono cuci bajunya? / Siapa yang Tono cuci baju=nya? / Who COMP Tono wash clothes=POSS / Lit.: 'Whose clothes did Tono wash of?' / 'Whose clothes did Tono wash?'	Siapa yang Tono cuci baju?	Must have =*nya* with WH-movement
WH-movement with adjunct and coordinate islands	Adjunct islands	Tono pulang setelah siapa menelepon dia? / Tono pulang setelah siapa men-telepon dia? / Tono return after who MEN-phone 3SG / 'Tono returned after who called him?'	Siapa yang Tono pulang setelah menelepon dia?	No WH-movement out of adjunct island possible
	Coordinate islands	Guru itu menghukum Tono dan siapa di sekolah minggu kemarin? / Guru itu men-hukum Tono dan siapa di sekolah minggu kemarin? / teacher DEM MEN-punish Tono and who in school week yesterday / 'That teacher punished Tono and who in school last week?'	Siapa yang guru itu menghukum Tono dan di sekolah minggu kemarin?	No WH-movement out of coordinate island possible

Filler-gap Dependencies

Object relatives and demonstratives	-	Mobil yang Tono tumpangi itu cepat sekali Mobil yang Tono tumpang-i itu cepat sekali car REL Tono ride-I DEM fast very 'The car that Tono was riding in was very fast.'	Mobil itu yang Tono tumpangi cepat sekali	Demonstrative *itu* cannot be inserted into the landing site of object in object relatives
Subject relatives and demonstratives	-	Wanita yang melarikan diri dari penjara itu seram sekali Wanita yang men-lari-kan diri dari penjara itu seram sekali woman REL MEN-run-KAN self from prison DEM scary very 'The woman who escaped from prison is very scary.'	Wanita yang itu melarikan diri dari penjara itu seram sekali	Demonstrative *itu* cannot be inserted in the base position of subject in subject relative
Topicalization with Left Branch Condition	-	Tini yang Tono temui ibunya Tini yang Tono temu-i ibu=nya Tini COMP Tono meet-I mother=NYA 'Tini is the one whose mother Tono met.'	Tini yang Tono temui ibu	Must have =*nya* with topicalization
NPIs and Negation — NPI licensing in-situ	-	Tono tidak sakit karena apa-apa Tono tidak sakit karena apa-apa Tono NEG sick because what-RED 'Tono was not sick because of anything.'	Tono sakit karena apa-apa	NPI must be licensed by c-commanding negation
NPI licensing with movement	-	Tono tidak tahu apa-apa Tono tidak tahu apa-apa Tono NEG know what-RED 'Tono does not know anything.'	Apa-apa yang Tono tidak tahu	NPI cannot be moved

Table 1: Examples of minimal pairs for each category with explanations