

# Deriving categorical and continuous properties of Javanese speech levels

Christopher Davis

AFLA 28  
McGill/NUS

May 27, 2021

# Speech Levels in Javanese

Javanese utterances can be sorted into one of three levels, called *ngoko*, *krama*, and *madya*, whose choice is conditioned by *status*, *age*, and *intimacy*, with the following canonical contexts of use:

- ▶ *Ngoko*: **low status** addressee, **not older** than the speaker, **intimate** relationship
- ▶ *Krama*: **high status** addressee, **older** than the speaker, **non-intimate** relationship
- ▶ *Madya*: a “halfway house” (Wolff & Poedjosoedarmo 1982) between *ngoko* and *krama*, canonically used in situations where the factors determining the choice of speech level are in conflict.

# Encoding of Speech Levels

Javanese utterances can be unambiguously assigned to one of the three levels (N, M, K) on a purely formal basis. The following example is from Clynes (1989):

(1) “Bu Siti has already eaten that one.”

**Krama** *Bu Siti sampun nedha ingkang menika.*

**Madya** *Bu Siti mpun nedha sing niku.*

**Ngoko** *Bu Siti wis mangan sing kuwi.*

Ms. Siti already eat REL that

Speech level is encoded through the choice between otherwise synonymous lexical alternants whose only difference is their (in)compatibility with particular speech levels.

# Encoding of Speech Levels

(1) “Bu Siti has already eaten that one.”

**Krama** *Bu Siti sampun nedha ingkang menika.*

**Ngoko** *Bu Siti wis mangan sing kuwi.*  
Ms. Siti already eat REL that

- ▶ The **blue** alternants are compatible with **Krama**, but not with **Ngoko**.
- ▶ The **red** alternants are compatible with **Ngoko**, but not with **Krama**.

# Encoding of Speech Levels

(1) “Bu Siti has already eaten that one.”

**Krama** *Bu Siti sampun nedha ingkang menika.*

**Madya** *Bu Siti mpun nedha sing niku.*

**Ngoko** *Bu Siti wis mangan sing kuwi.*

Ms. Siti already eat REL that

- ▶ **Madya** is characterized by a mixture of **Ngoko** and **Krama** forms. . .

# Encoding of Speech Levels

(1) “Bu Siti has already eaten that one.”

**Krama** *Bu Siti sampun nedha ingkang menika.*

**Madya** *Bu Siti mpun nedha sing niku.*

**Ngoko** *Bu Siti wis mangan sing kuwi.*

Ms. Siti already eat REL that

- ▶ **Madya** is characterized by a *mixture* of **Ngoko** and **Krama** forms,
- ▶ and some forms that are only compatible with **Madya**.

# Encoding of Speech Levels

(1) “Bu Siti has already eaten that one.”

**Krama** *Bu Siti sampun nedha ingkang menika.*

**Madya** *Bu Siti mpun nedha sing niku.*

**Ngoko** *Bu Siti wis mangan sing kuwi.*

Ms. Siti already eat REL that

- ▶ **Madya** is characterized by a *mixture* of **Ngoko** and **Krama** forms,
- ▶ and some forms that are **only** compatible with **Madya**.
- ▶ Similarly, some forms are **only** compatible with **Ngoko**...

# Encoding of Speech Levels

(1) “Bu Siti has already eaten that one.”

<b>Krama</b>	<i>Bu</i>	<i>Siti</i>	<u>sampun</u>	<i>nedha</i>	<u>ingkang</u>	<u>menika</u> .
<b>Madya</b>	<i>Bu</i>	<i>Siti</i>	mpun	<i>nedha</i>	<i>sing</i>	<b>niku</b> .
<b>Ngoko</b>	<i>Bu</i>	<i>Siti</i>	wis	<b>mangan</b>	<i>sing</i>	<b>kuwi</b> .
	Ms.	Siti	already	eat	REL	that

- ▶ **Madya** is characterized by a *mixture* of **Ngoko** and **Krama** forms,
- ▶ and some forms that are **only** compatible with **Madya**.
- ▶ Similarly, some forms are **only** compatible with **Ngoko**,
- ▶ while others are only compatible with **Krama**.



# Analysis: Lexical Classes and Speech Levels

(1) “Bu Siti has already eaten that one.”

**Krama** *Bu Siti sampun nedha ingkang menika.*

**Madya** *Bu Siti mpun nedha sing niku.*

**Ngoko** *Bu Siti wis mangan sing kuwi.*

Ms. Siti already eat REL that

- ▶ Words like **sampun** can only be used in Krama
- ▶ Words like *nedha* can be used in both Madya and Krama
- ▶ Words like **mpun** can only be used in Madya
- ▶ Words like *sing* can be used in both Ngoko and Madya
- ▶ Words like **wis** can only be used in Ngoko

# Analysis: Lexical Classes and Speech Levels

(1) “Bu Siti has already eaten that one.”

<b>Krama</b>	<i>Bu</i>	<i>Siti</i>	<b>sampun</b>	<i>nedha</i>	<b>ingkang</b>	<b>menika.</b>
<b>Madya</b>	<i>Bu</i>	<i>Siti</i>	<b>mpun</b>	<i>nedha</i>	<i>sing</i>	<b>niku.</b>
<b>Ngoko</b>	<i>Bu</i>	<i>Siti</i>	<b>wis</b>	<b>mangan</b>	<i>sing</i>	<b>kuwi.</b>
	Ms.	Siti	already	eat	REL	that

- ▶ Words like **sampun**:  $[-N, +K]$
- ▶ Words like *nedha*:  $[ \quad +K]$
- ▶ Words like **mpun**:  $[+N, +K]$
- ▶ Words like *sing*:  $[+N \quad ]$
- ▶ Words like **wis**:  $[+N, -K]$

# Linking Lexical Classes and Speech Levels

Speech Levels:

- ▶ **Krama** is signaled by  $-N, +K$
- ▶ **Madya** is signaled by  $+N, +K$
- ▶ **Ngoko** is signaled by  $+N, -K$

Combinatoric Constraint: Feature values must be consistent.

- ▶ A single utterance cannot contain both  $+N$  and  $-N$  items.
- ▶ A single utterance cannot contain both  $+K$  and  $-K$  items.
- ▶ Another way of looking at this: An utterance must belong to one of the three speech levels.

# Linking Lexical Classes and Speech Levels

(1) “Bu Siti has already eaten that one.”

**Krama** *Bu Siti sampun nedha ingkang menika.*

**Madya** *Bu Siti mpun nedha sing niku.*

**Ngoko** *Bu Siti wis mangan sing kuwi.*

Ms. Siti already eat REL that

		Ngoko	Madya	Krama
<b>sampun</b>	$[-N, +K]$	×	×	
<i>nedha</i>	$[ \quad +K]$	×		
<b>mpun</b>	$[+N, +K]$	×		×
<i>sing</i>	$[+N \quad ]$			×
<b>wis</b>	$[+N, -K]$		×	×

# Paradigm 1

(1)

P1

**Krama** *Bu Siti* sampun *nedha* **ingkang** *menika.*  
**Madya** *Bu Siti* mpun *nedha* *sing* **niku.**  
**Ngoko** *Bu Siti* wis **mangan** *sing* **kuwi.**  
 Ms. Siti already eat REL that

		Ngoko	Madya	Krama
<b>sampun</b>	$[-N, +K]$	×	×	
<b>mpun</b>	$[+N, +K]$	×		×
<b>wis</b>	$[+N, -K]$		×	×

# Paradigm 1

(1)

			<b>P1</b>			<b>P1</b>
<b>Krama</b>	<i>Bu</i>	<i>Siti</i>	<b>sampun</b>	<i>nedha</i>	<b>ingkang</b>	<u><b>menika</b></u> .
<b>Madya</b>	<i>Bu</i>	<i>Siti</i>	<b>mpun</b>	<i>nedha</i>	<i>sing</i>	<u><b>niku</b></u> .
<b>Ngoko</b>	<i>Bu</i>	<i>Siti</i>	<b>wis</b>	<b>mangan</b>	<i>sing</i>	<u><b>kuwi</b></u> .
	Ms.	Siti	already	eat	REL	that

		Ngoko	Madya	Krama
<b>menika</b>	$[-N, +K]$	×	×	
<b>niku</b>	$[+N, +K]$	×		×
<b>kuwi</b>	$[+N, -K]$		×	×

# Paradigm 2

(1)

			P1	P2		P1
<b>Krama</b>	<i>Bu</i>	<i>Siti</i>	<b>sampun</b>	<u><i>nedha</i></u>	<b>ingkang</b>	<b>menika.</b>
<b>Madya</b>	<i>Bu</i>	<i>Siti</i>	<b>mpun</b>	<u><i>nedha</i></u>	<i>sing</i>	<b>niku.</b>
<b>Ngoko</b>	<i>Bu</i>	<i>Siti</i>	<b>wis</b>	<u><b>mangan</b></u>	<i>sing</i>	<b>kuwi.</b>
	Ms.	Siti	already	eat	REL	that

		Ngoko	Madya	Krama
<i>nedha</i>	[ +K ]	×		
<b>mangan</b>	[ +N, -K ]		×	×

# Paradigm 3

(1)

			P1	P2	P3	P1
<b>Krama</b>	<i>Bu</i>	<i>Siti</i>	<b>sampun</b>	<i>nedha</i>	<u><b>ingkang</b></u>	<b>menika.</b>
<b>Madya</b>	<i>Bu</i>	<i>Siti</i>	<b>mpun</b>	<i>nedha</i>	<u><i>sing</i></u>	<b>niku.</b>
<b>Ngoko</b>	<i>Bu</i>	<i>Siti</i>	<b>wis</b>	<b>mangan</b>	<u><i>sing</i></u>	<b>kuwi.</b>
	Ms.	Siti	already	eat	REL	that

		Ngoko	Madya	Krama
<b>ingkang</b>	$[-N, +K]$	×	×	
<i>sing</i>	$[+N]$			×



# Continuous Properties of the Madya Level

- ▶ Clynes (p.45), citing Uhlenbeck (1970:452), W&P:17, and Errington (1985:107):  
*“In the intermediate madya style, the degree of relative “formality” or “distance” is directly dependent on the relative proportions of ngoko and krama (general lexis) items used.”*
- ▶ W&P (p.17):  
*“Madyô is not a set of fixed forms, but is rather a cline rising from a level very close to Ngoko up to a level very close to Krômô. The height of the M level depends on the mixture of Ngoko and Krômô. . . The greater the percentage of N forms . . . the lower the M level.”*

# Continuous Properties of the Madya Level

- ▶ Speech levels are categorically determined, but Madya level allows for both [+K] and [+N] lexical items (of which there are many, something like 500 each).
- ▶ Madya speech level utterances sit on a “more krama-like” to “more ngoko-like” continuum.
- ▶ Analytic intuition: The position of a particular Madya-level utterance along this continuum is (partly) a function of the proportion of [+K] and [+N] lexical items.

# Overview of Data

The rest of this talk will be based on data from Wolff & Poedjosoedarmo 1982 (henceforth W&P) which were created as follows:

- ▶ A large natural corpus of Javanese dialogs was recorded.
- ▶ Native speakers then assigned utterances from the corpus into speech levels (Ngoko, Madya, Krama).
- ▶ They assigned Madya-level utterances to one of three sub-levels: Low Madya (LM), Mid Madya (M), and High Madya (HM).
- ▶ The occurrence or non-occurrence of particular lexical items across these levels was determined, and is presented in a series of tables (pp. 30–35).
- ▶ The data from these tables form the basis for the rest of this talk.

# Overview of Data: Categorical vs Continuous Distinctions

W&P (p.29): “How did we determine these speech levels? First, we took a portion of our materials and had native speakers assign a speech level to each utterance: K (Krômô), MT (for Madyô Tinggi, High Madyô), M (Madyô, not high or low), MR (for Madyô Rendah, Low Madyô), and N (Ngoko). We tested these identifications and received nearly 100 percent agreement on assignment into three categories, K, M, and N; but the MT, M, MR distinction was impossible to make consistently, as we ourselves had been unable to specify what the differences were. Thus, the informants distinguished the various kinds of Madyô impressionistically.”

# Paradigms 1, 2, and 3: Observed Distributions

		Speech Level				
		N	LM	M	HM	K
'already'						
sampon	$[-N, +K]$					█
ampon	$[+N, +K]$		█	█	█	
wés	$[+N, -K]$	█				
'who'						
sinten	$[ \quad +K]$		█	█	█	█
sôpô	$[+N, -K]$	█				
REL						
éngkang	$[-N, +K]$					█
séng	$[+N \quad ]$	█	█	█	█	

# Paradigms 1, 2, and 3: No Effect on Madya Sublevel

		Speech Level				
		N	LM	M	HM	K
'already'						
sampon	$[-N, +K]$					■
ampon	$[+N, +K]$		■	■	■	
wés	$[+N, -K]$	■				
'who'						
sinten	$[ \quad +K]$		■	■	■	■
sôpô	$[+N, -K]$	■				
REL						
éngkang	$[-N, +K]$					■
séng	$[+N \quad ]$	■	■	■	■	

# Paradigm 4: Overlap and competition

Table: Paradigm 2

		Speech Level				
		N	LM	M	HM	K
'who'						
sinten	[ +K ]		████████████████████			████
sôpô	[ +N, -K ]	████				

# Paradigm 4: Overlap and competition

Table: Paradigm 2

'who'		Speech Level				
		N	LM	M	HM	K
sinten	[ + K ]		█			█
sôpô	[ +N, - K ]	█				

Table: Paradigm 4

'how'		Speech Level				
		N	LM	M	HM	K
kadôs pundi	[ + K ]		×	×	█	█
(ke)pripon	[ +N, + K ]		█		×	
(ke)priyé	[ +N, - K ]	█				



# Paradigm 4: Variation in patterns

## Pattern 1

		Speech Level				
		N	LM	M	HM	K
'how'						
kadôš pundi	[ + K]		×	×		
(ke)ripon	[+N, + K]				×	
(ke)riyé	[+N, - K]					

## Pattern 2

		N	LM	M	HM	K
'from'						
sakéng	[ + K]					
(se)kéng	[+N, + K]				×	
sekô/sôkô	[+N, - K]					

# Paradigm 5: Same thing, other direction

Table: Paradigm 3

REL		Speech Level				K
		N	LM	M	HM	
éngkang	$[-N, +K]$					
séng	$[+N \quad ]$					

# Paradigm 5: Same thing, other direction

Table: Paradigm 3

REL		Speech Level				
		N	LM	M	HM	K
éngkang	$[-N, +K]$					
séng	$[+N]$					

Table: Paradigm 5

'don't'		Speech Level				
		N	LM	M	HM	K
sampon	$[-N, +K]$					
ampon	$[+N, +K]$		×	??		
ôjô	$[+N]$			??	×	

# Paradigm 6: Three-way competition

Table: Paradigm 6

'place'	Speech Level				
	N	LM	M	HM	K
panggènan	[ + K ]	×	×		
nggèn	[ +N, + K ]			×	
nggôn	[ +N ]			×	

# Interim Summary

## Non-Competitive Paradigms

P1	P2	P3
$[-N, +K]$	$[ \quad +K]$	$[-N, +K]$
$[+N, +K]$	$[+N, -K]$	$[+N \quad ]$
$[+N, -K]$		

## Competitive Paradigms

P4	P5	P6
$[ \quad +K]$	$[-N, +K]$	$[ \quad +K]$
$[+N, +K]$	$[+N, +K]$	$[+N, +K]$
$[+N, -K]$	$[+N \quad ]$	$[+N \quad ]$

# Lexical Classes by Number

Table: Lexemes by class (From Clynes)

	number (approx.)	% of lexicon
ngoko	580	3
krama	580	3
madya	30	< 0.2%
deferential	210	1
neutral	c.20,000	93

- ▶ About 30 each of the *krama* and *ngoko* lexemes are “style-markers”, that is, either  $[+K, -N]$  or  $[-K, +N]$ .
- ▶ The rest are “general lexis”, that is, either  $[+K]$  or  $[+N]$ .
- ▶ The great majority of lexical alternant sets thus involve a two-way alternation between a  $[+K]$  and  $[+N]$  forms.
- ▶ I call this “Paradigm 0”.

# Paradigm 0, Pattern 1

Table: Pattern 1a (partial list of 47 sets in W&P)

		Speech Level				
		N	LM	M	HM	K
'house'						
griyô	[ +K ]		×			
omah	[ +N ]			×	×	

Table: Pattern 1b (12 sets in W&P)

		Speech Level				
		N	LM	M	HM	K
'child'						
laré	[ +K ]		%			
bocah	[ +N ]			×	×	

# Paradigm 0, Pattern 2

Table: Pattern 2 (11 sets in W&P)

'as'	Speech Level				
	N	LM	M	HM	K
kadôs	[ +K ]	×	×		
kôyô	[ +N ]			×	



# Comparison of Paradigm 0 Patterns

		Speech Level				
		N	LM	M	HM	K
Pattern 1a						
griyô	[ + K ]		×			
omah	[ + N ]			×	×	
Pattern 1b						
laré	[ + K ]		%			
bocah	[ + N ]			×	×	
Pattern 2						
kadô	[ + K ]		×	×		
kôyô	[ + N ]				×	

# Diachronic Speculations

Generalization: Paradigm 0 (two-member sets whose members are marked as [+K] and [+N]) accounts for the vast majority of alternant sets (something like 500 sets, 1000 lexical items, using the numbers in Clynes).

- ▶ Speculation 1: Paradigm 0 is both synchronically and diachronically the core of the speech level system.
- ▶ Speculation 2: The other paradigms evolved from Paradigm 0

# Diachronic Speculations

Speculation 1: Paradigm 0 is both synchronically and diachronically the core of the system.

- ▶ The original speech level system would have had only Paradigm 0 alternant sets (two-way contrast of  $[+K]$  and  $[+N]$  alternants).
- ▶ The *krama* and *ngoko* speech levels would have been characterized by utterances that exclusively used either  $[+K]$  or  $[+N]$  forms.
- ▶ The *madya* level would have been characterized by utterances that mixed  $[+K]$  and  $[+N]$  forms.
- ▶ Note: This system would not have had any categorical restrictions on combinations of forms.

# Diachronic Speculations

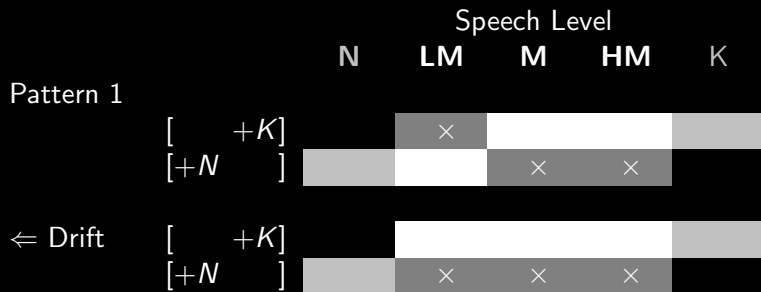
Speculation 2: The other paradigms evolved from Paradigm 0 by the following processes:

1. Competition between  $[+K]$  and  $[+N]$  in the *madya* level led, in some cases, to an evolution from either:
  - ▶  $[ \quad +K ]$  to  $[-N, +K]$ , or
  - ▶  $[+N \quad ]$  to  $[+N, -K]$
2. Alternative *krama* forms with differing levels of formality led to the creation of *madya* forms.

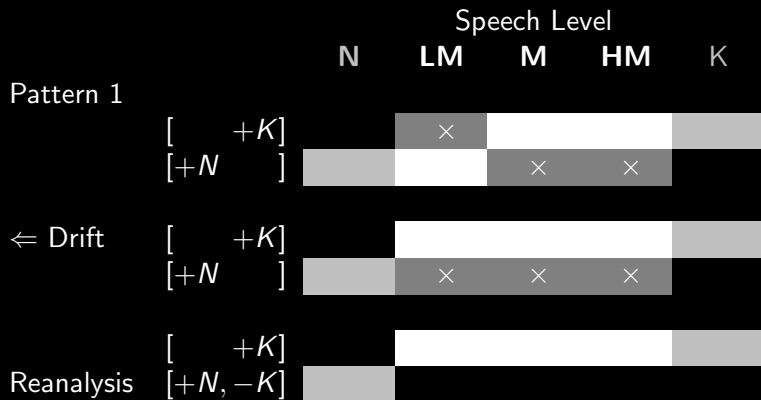
# Paradigm 0 → Paradigm 2



# Paradigm 0 → Paradigm 2



# Paradigm 0 $\rightarrow$ Paradigm 2



# Paradigm 0 → Paradigm 3

Pattern 2

		Speech Level				
		N	LM	M	HM	K
[	+K]		×	×		
[+N	]				×	



# Paradigm 0 → Paradigm 3

		Speech Level				
		N	LM	M	HM	K
Pattern 2	$\begin{bmatrix} +K \\ +N \end{bmatrix}$		×	×		
					×	
⇒ Drift	$\begin{bmatrix} +K \\ +N \end{bmatrix}$		×	×	×	

# Paradigm 0 → Paradigm 3

		Speech Level				
		N	LM	M	HM	K
Pattern 2	$\begin{bmatrix} +K \\ +N \end{bmatrix}$		×	×		
⇒ Drift	$\begin{bmatrix} +K \\ +N \end{bmatrix}$		×	×	×	
Reanalysis	$\begin{bmatrix} -N, +K \\ +N \end{bmatrix}$					

## Speculation 2: The Krama to Madya pipeline

- ▶ Many (not all) *madya* forms derive diachronically from a reduced, “casual” variant of the associated *krama* form.
- ▶ Some of these you have already seen:

# Speculation 2: The Krama to Madya pipeline

- ▶ Many (not all) *madya* forms derive diachronically from a reduced, “casual” variant of the associated *krama* form.
- ▶ Some of these you have already seen:

		Speech Level				
		N	LM	M	HM	K
'already' (P1)						
sampon	$[-N, +K]$					█
ampon	$[+N, +K]$		█	█	█	
wés	$[+N, -K]$	█				
'from' (P4)						
sakéng	$[ \quad +K]$		█	█	█	█
(se)kéng	$[+N, +K]$		█	█	█	█
sekô/sôkô	$[+N, -K]$	█			█	
'don't' (P5)						
sampon	$[-N, +K]$					█
ampon	$[+N, +K]$		█	??	█	
ôjô	$[+N \quad ]$	█		??	█	

# Speculation 2: The Krama to Madya pipeline

		Speech Level				
		N	LM	M	HM	K
<b>'already' (P1)</b>						
sampon	$[-N, +K]$					█
ampon	$[+N, +K]$		█	█	█	
wés	$[+N, -K]$	█				
<b>'from' (P4)</b>						
sakéng	$[ \quad +K]$		█	█	█	█
(se)kéng	$[+N, +K]$		█	█	█ ×	
sekô/sôkô	$[+N, -K]$	█				
<b>'don't' (P5)</b>						
sampon	$[-N, +K]$					█
ampon	$[+N, +K]$		█ ×	█ ??	█	
ôjô	$[+N \quad ]$	█	█	█ ??	█ ×	

## Speculation 2: The Krama to Madya pipeline

- ▶ Assume an original Paradigm 0 set.
- ▶ Reduced forms of the [+K] alternant are used in more casual contexts (eg, Madya speech level contexts).
- ▶ This drives a reanalysis of the reduced form as a dedicated Madya form.
- ▶ This would derive something like our Paradigm 6:

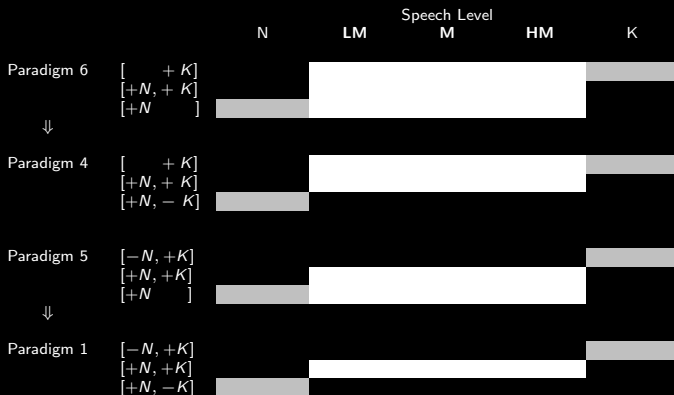
Table: Paradigm 6

		Speech Level				
'place'		N	LM	M	HM	K
panggènan	[ +K ]		×	×		
nggèn	[ +N, +K ]				×	
nggôn	[ +N ]				×	

# Speculation 2: The Krama to Madya Pipeline

Further competition could trigger further reanalysis:

1. Paradigm 6  $\Rightarrow$  Paradigm 4 or 5
2. Paradigm 4 or 5  $\Rightarrow$  Paradigm 1



# Speculation 2: The Krama to Madya Pipeline

Further competition could trigger further reanalysis:

1. Paradigm 6  $\Rightarrow$  Paradigm 4 or 5
2. Paradigm 4 or 5  $\Rightarrow$  Paradigm 1

		Speech Level				
		N	LM	M	HM	K
'place' (P6)						
panggènan	[ +K ]		x	x		
nggèn	[ +N, +K ]				x	
nggôn	[ +N ]				x	
'from' (P4)						
sakéng	[ +K ]					
(se)kéng	[ +N, +K ]				x	
sekô/sôkô	[ +N, -K ]					
'don't' (P5)						
sampon	[ -N, +K ]					
ampon	[ +N, +K ]		x	??		
ôjô	[ +N ]			??	x	
'already' (P1)						
sampon	[ -N, +K ]					
ampon	[ +N, +K ]					
wés	[ +N, -K ]					



# Conclusion

- ▶ The paradigms discussed here cover most if not all of the patterns found in the data of Wolff and Poedjosoedarmo.
- ▶ There are a handful of cases that present complications; these will be discussed in more detail in the paper (hint hint).
- ▶ The “current” (as of the late 20th century) speech level system is fundamentally categorical in nature.
- ▶ But most alternant sets give rise to a choice in the Madya level that is not categorically determined.
- ▶ The choice in such cases determines how “Ngoko-like” or “Krama-like” the utterance is.
- ▶ I speculate that such competition gave rise to the five lexical classes and seven paradigms from an original system with only two lexical classes and one paradigm.

Thank you!

Questions and comments please!