Investigating Colloquial Indonesian: Nasal assimilation differences in Betawi Malay and Jakarta Indonesian
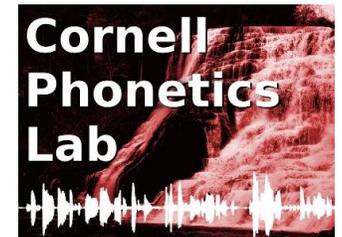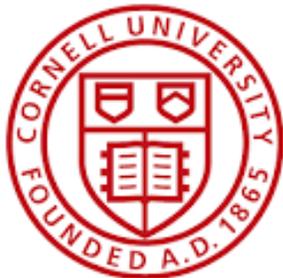
Ferdinan Okki Kurniawan

fok3@cornell.edu

13-ICAL

Department of Linguistics

Cornell University

Purpose of the study

- Investigates nasal assimilation produced by Betawi Malay (BM) and Jakarta Indonesian (JI) speakers.

- Provides a twofold contribution:
  - Rather than relying on impressionistic observation, data in this study are drawn from a naturalistic spoken corpus (Gil and Tadmor 2014) and a production task.

  - Finds that the speakers of JI produce variation in the nasal assimilation at the prefix-root boundary.
    - Inter-speaker variation
    - Intra-speaker variation

Outline of the talk

I.   Previous Studies

II. An overview of Standard Indonesian (SI), Jakarta Indonesian (JI) and Betawi Malay (BM)

III. Nasal Prefix (N-prefix) in SI

IV. Nasal Prefix (N-prefix) in JI and BM

V.   Reviewing Zuraw's (2010) model of Lexical Variation

VI. Results from corpus

VII. Results from production task

VIII. Summary

I. Previous Studies

Previous studies

- Pater (1999, 2001) investigates formal driving force of nasal assimilation in Indonesian within the Optimality Theory framework (OT; Prince and Smolensky 1993, 2004).

- This pattern was first described systematically by Lapoliwa (1981) within a generative framework.

- All previous studies were devoted to the standard variety of Indonesian spoken in formal contexts (standard Indonesian (SI)).

- Less attention, however, has been given to the pattern of nasal assimilation in a more colloquial variety of Indonesian as a casual everyday language spoken in Jakarta (Jakarta Indonesian (JI))

II. An overview of Standard Indonesian (SI), Jakarta Indonesian (JI) and Betawi Malay (BM)

An overview of Standard Indonesian (SI), Jakarta Indonesian (JI) and Betawi Malay (BM)

http://www.lonelyplanet.com/maps/asia/indonesia/

- JI: new emerging colloquial variety, spoken in Jakarta, the capital city of Indonesia (Sneddon 2006)

- SI: national and official language of Indonesia

- Betawi Malay (BM): older variety of Malay (emerging around 18th century) (Wallace 1976, among others)

- Diglossic situation forming dialect continuum:

BM -- JI -- SI

An overview of Standard Indonesian (SI), Jakarta Indonesian (JI) and Betawi Malay (BM)

- Standard Indonesian:
  - SI is not spoken on daily basis and only spoken in formal situations.
  - SI is taught formally at school and most children have limited or no exposure to this standard variety until they start to go to school (Sneddon 2006).

- Jakarta Indonesian:
  - As a vernacular spoken at home, JI is acquired by children in Jakarta as their first language.
  - As a vernacular, JI has emerged since World War II, followed by huge influx of migrants into the capital city. The second generation of these migrant families have been forming this new linguistic variety called Jakarta Indonesian (modern Jakarta Malay) (Wallace 1976).
  - JI is spoken widely by more educated speakers associated with higher socio-economic status in Jakarta (Sneddon 2006).

- Betawi Malay:
  - Nowadays, BM is spoken by a small minority group in Jakarta.
  - BM is the vernacular spoken by inhabitants of 'old kampungs' in Jakarta, limited to Betawi communities (Sneddon 2006).

An overview of Standard Indonesian (SI), Jakarta Indonesian (JI) and Betawi Malay (BM)

- The phonology of BM, JI and SI are different in some important respects but are mutually intelligible.

- <u>Differences in the patterns of use of nasal assimilation</u> show evidence of this.

- This study, however, *does not* aim to address the socio-phonological differences between JI and BM. Rather, this investigation uses these two socio-ethnic groups to map the variations of the nasal assimilation.

III. Nasal Prefix (N-prefix) in SI

N-prefix in Standard Indonesian

- In SI, a nasal in the coda position of the prefix /məN-/ alternates when it patterns with root-initial consonants.

  - This study focus on obstruents

- It is assumed widely that the underlying nasal coda of the prefix is a placeless nasal which is symbolized with *N* in *məN-* above (following Muhadjir 1981, Sneddon 1996 and 2006, Pater 1999 among others).

Standard Indonesian: N-prefix + root-initial voiced consonants

- N-prefix + root-initial voiced consonants :
  (a) /məN+bəli/   məmbəli   'to buy'
  (b) /məN+guntiŋ/   məŋguntiŋ   'to cut with scissors'
  (c) /məN+dapat/   mendapat   'to get'
  (d) /məN+dʑawab/   məɲdʑawap   'to answer'

- assimilation
- homorganic clusters occurs

Standard Indonesian: N-prefix + root-initial voiceless consonants

- ## N-prefix + root-initial voiceless consonants:

  (a) /məN+pilih/   məmilih   'to choose, to vote'

  (b) /məN+tulis/   mənulis   'to write'

  (c) /məN+kasih/   məŋasih   'to give'

  (d) /məN+sapu/   məɲapu   'to sweep'

Rule based account: substitution (assimilation followed by deletion)

  (e) /məN+tɕari/   məɲtɕari   'to seek'

Does not undergo deletion

JI, BM and SI Nasal Prefix (N-prefix)

- SI: məN-

- JI and BM: N-

IV. N-prefix in JI and BM

N-prefix in JI and BM

- The N-prefix in JI and BM is cognate with the N-prefix *məN-* in SI.

- The description in the next slides is based on my observations and coincides closely with the patterns described by Ikranagara (1980), Muhadjir (1981) and Sneddon (2006).

JI and BM: N-prefix + root-initial voiceless consonants

N-prefix + root-initial voiceless consonants

    (a) /N+pilih/   milih   'to choose, to vote'
    (b) /N+tulis/   nulis   'to write'
    (c) /N+kasih/   ŋasih   'to give'
    (d)/N+sapu/   ɲapu   'to sweep'

- Substitution: assimilation followed by deletion

    - (e) /N+tɕari/   ɲari   'to seek'
- undergoes deletion

N-prefix in SI, JI and BM

- SI : /məN+bəli/ məmbəli 'to buy'

- JI and BM : /N+bəli/ mbəli 'to buy'

- JI and BM : /N+bəli/ ŋəbəli 'to buy'

- JI and BM : /N+bəli/ mbəli ~ ŋəbəli 'to buy'

JI and BM: N-prefix +root-initial voiced consonants

- Either schwa epenthesis or homorganic clusters occurs in root-initial voiced consonants:

(a) /N+beli/   ŋəbəli ~ mbəli   'to buy'
(b) /N+dapet/   ŋədapət ~ ndapət   'to get'
(c) /N+guntiŋ/   ŋəguntiŋ ~ ŋguntiŋ   'to cut with scissors'
(d) /N+dʑawab/   ŋədʑwap ~ ndʑawap   'to answer'

How do we account for such variations?

- One of the key questions to be addressed is what conditions this variation and why.

- One of the central goals of this study is to discuss this pattern of variation.

Before looking more closely at the patterns of variation, it is useful to look at a similar pattern observed in Tagalog, a western Austronesian language, as discussed by Zuraw (2010).

V. Reviewing Zuraw's (2010) model of Lexical Variation

Lexical variation model (Zuraw 2010) of nasal assimilation in Tagalog data

- **Tagalog data (zuraw 2010):**

  - /paŋ + **p**oʔók/   pam-**p**oʔók   'local'
  - /paŋ-RED + **p**ighatiʔ/   pa-mi-**m**ighatiʔ   'being in grief'
  - (generated to t, k, s, ʔ- root initials)

  - /maŋ-RED + **b**igkás/   ma**m**-**b**i-**b**igkás   'reciter'
  - /maŋ + mag-**b**igaj 'to give'/   ma-**m**igaj   'to distribute'
  - (generated to d, g- root initials)

Lexical variation model (Zuraw 2010)

- Types of variations, Zuraw (2010):

  - **lexical variation**: each word is considered to have just one behavior and it applies across lexicon

  - **free variation**: the same speaker can produce two different output

  - free variation is often resulted from **stylistic variable** which is affected by social factors such as degree of formality, dialect, etc.

Zuraw's (2010) study on the lexical variation model of Tagalog nasal substitution

- Her study proposed three type classifications:

    - the first type is the words that are **lexicalized in the form of nasal substitution**;

    - the second type includes the words that are **lexicalized but do not undergo nasal substitution**;

    - the last type are the words that are **not yet lexicalized**.

Is variation in nasal assimilation and substitution in JI native words and SI borrowings also due to lexical variation?

Naturalistic spoken corpus and production task investigation

VI. Results from corpus

Description of corpus

- The corpus of Betawi-Jakarta project (Gil and Tadmor 2004-2014)
- Collected in Jakarta
- Duration of each recording: 30 to 60 minutes
- A total of 75,709 utterances transcribed into a computerized database
- Searchable by orthography, IPA, morphology
- Transcribed by trained linguists

Hypotheses for JI corpus investigation

- Hypothesis 1: The patterns of variation in the output occur due inter-speaker variation
- Hypothesis 2: The patterns of variation in the output occur due to intra-speaker variation
- Hypothesis 3: The patterns of variation in the output occur due to lexical variation
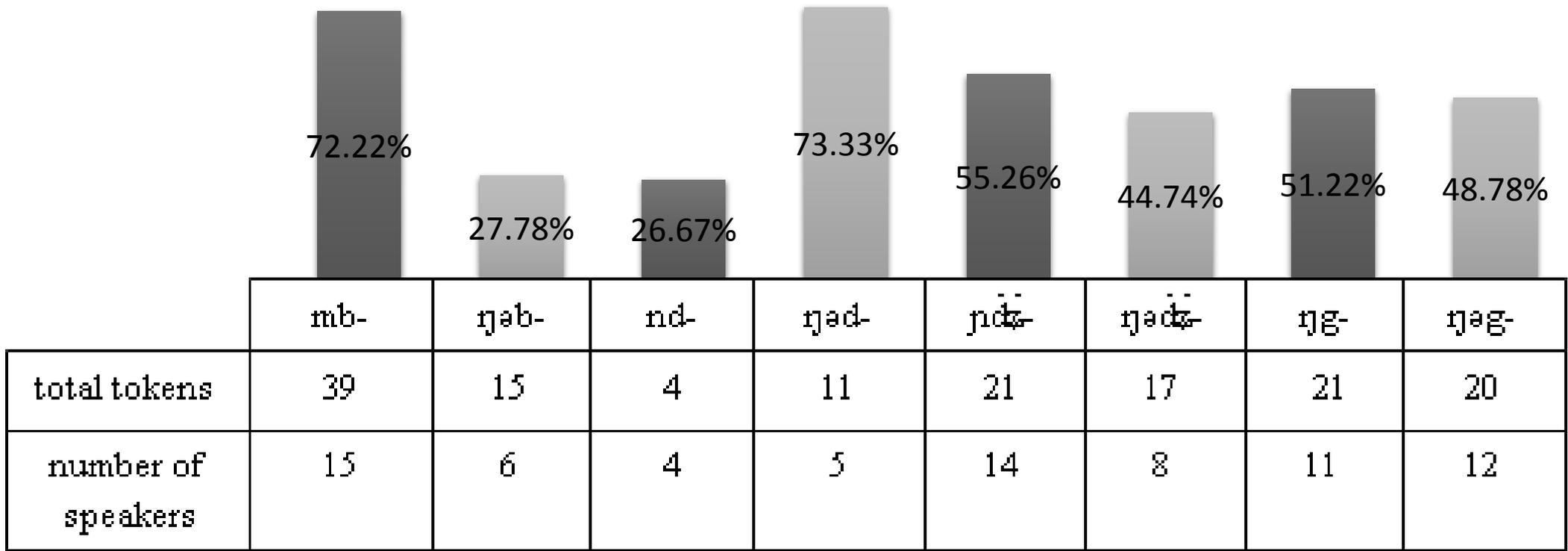
Inter-speaker variation

Findings: inter-speaker variation/dialect (hypothesis 1)

- The speakers in the corpus are divided into two main groups:
  - Speakers of BM and
  - Speakers of JI.

- Based on speakers' description provided in JFS metadata.

- Speakers of BM: the speakers of Betawi ethnicity

- Speakers of JI: those who have ethnic background other than Betawi but were born and grew up in Jakarta and their parents are not of Betawi ethnicity

- Wallace (1976) termed BM is 'traditional Betawi Malay' while JI is 'modern Betawi Malay'
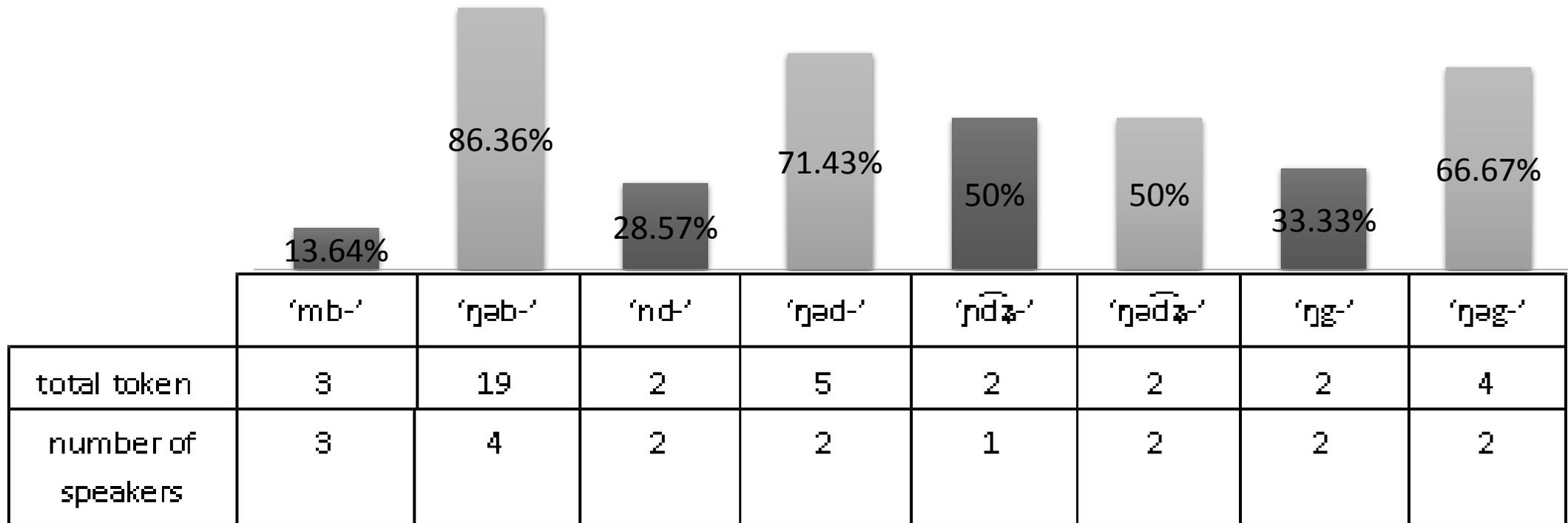
Corpus search

- To test hypothesis 1
- To observe whether dialect groups exhibit certain patterns of N-prefix:
  - no-schwa form (homorganic clusters) [mb-], [nd-], [ŋg-], [ndʑ-] or

  - schwa form: [ŋəb], [ŋəg-], [ŋəd-], [ŋədʑ-].

Total tokens by BM speakers

	mb-	ŋəb-	nd-	ŋəd-	ɲʥ-	ɲəʥ-	ŋg-	ŋəg-
	72.22%	27.78%	26.67%	73.33%	55.26%	44.74%	51.22%	48.78%
total tokens	39	15	4	11	21	17	21	20
number of speakers	15	6	4	5	14	8	11	12

Total tokens by JI speakers

	'mb-'	'ŋab-'	'nd-'	'ŋad-'	'ɲd͡ʑ-'	'ŋad͡ʑ-'	'ŋg-'	'ŋag-'
total token	3	19	2	5	2	2	2	4
number of speakers	3	4	2	2	1	2	2	2

Bar chart percentages: 13.64%, 86.36%, 28.57%, 71.43%, 50%, 50%, 33.33%, 66.67%

Percentage of nasal assimilation (homorganic clusters) across place of articulation

Labial		Alveolar		Palatal		Velar	
BM	JI	BM	JI	BM	JI	BM	JI
72,22%	**13.64%**	26.67%	28.57%	55.26%	50%	**51.22%**	**33.33%**

Dialect groups: BM vs JI

- BM Malay speakers produce higher token for homorganic clusters, unless for [ŋəd-] and [nd-] where [ŋəd-] is higher than [nd-]

- JI speakers produce higher token for epenthesis

Intra-speaker variation

Intra-speaker variation (Hypothesis 2)

- Examples:

Speakers	Underlying representation	Output 1	Output 2
1	/N+gajəm/ 'to eat'	ŋəgajəm (1)	ŋgajəm (1)
2	/N+daptar/ 'to register'	ŋədaptar (1)	ndaptar (1)
3	/N+gəletak/ 'to lie down'	ŋəgəletak (1)	ŋgəletak (1)
4	/N+batɕa/ 'to read'	ŋəbatɕa (2)	mbatɕa (2)
5	/N+dʑual/ 'to sell'	ŋədʑual (1)	ndʑual (2)
6	/N+dəŋər+IN/ 'to listen'	ŋədəŋərin (1)	ndəŋərin (1)

- The number in the parenthesis indicates the numbers of tokens.
- the same speaker sometimes produces both forms even for the same lexical item.

No lexical variation

- These results are different from what Zuraw (2010) found based on her study in the written and dictionary corpus of Tagalog.

- The general lexical trend from the Tagalog written corpus is that the lexical items are already lexicalized or listed as a single lexical item in the Tagalog speakers' mind and while the same initial sound might show variation, there is no variation between speakers or within speaker for particular lexical items.

- In this current study, intra-speaker variation occurs in Betawi-Jakarta corpus where both forms can be uttered in spontaneous speech by the same speaker.

Summary from the corpus findings

- Hypothesis 1: the patterns of variation in the output occur due to inter-speaker variations)
- Hypothesis 2: the patterns of variation in the output occur due to intra-speaker variation
- Hypothesis 3: no lexical variation found

VII. Results from production task - JI

Production task – proposed hypothesis

- We can formulate this into H_1: ^{n}epenthetic schwa > ^{n}homorganic clusters.

- The null hypothesis for this study would be H_0 : ^{n}epenthetic schwa $\equiv$ ^{n}homorganic clusters.

Production task – JI speakers

- Test items:
  - 48 test items (12 words of b, d, g, dʑ-initial respectively)
  - Words are taken from the corpus
  - Embedded in different sentences

- Subjects:
  - The production task was conducted with 7 Indonesian speakers in Ithaca, New York, USA.
    - They are Cornell University graduate students, their spouses or faculty.
    - The age range of participants is between 25-45 years old.
    - Educated speakers of Indonesian

Production task - procedure

- Procedure:
    - Instruction are given at the begining of the task
    - 6 pre-training items:
        - Subjects listen to male voices that produce passive sentence in DI-prefix form, then followed by female voice that produce active sentence in the N-prefix form
        - p, t, k-initial are used

    - 9 training items:
        - p, t, k-initial are used
        - Subject produce N-prefix

    - Test items:
        - 48 test items are embedded in two different sentences;
        - distracter are placed in every 6 test items
        - 48 test items (12 words of b, d, g, dʑ-initial respectively)
        - Words are taken from the corpus
- The subjects are provided with the passive form sentences with DI-prefix construction and then are asked to produce the active form where N-prefix is expected to occur.

Production task – examples of test items

- The sentences and their topics are composed in a colloquial style and everyday situations to avoid participants producing SI forms.

- The example below is some of the test items in the production task:

T1: warung-warung   liar   itu   **di-bongkar-in**   satpol   PP   tadi   malem

   RED-stalls   illegal   that   PASS-take.apart-IN   police.unit   PP   earlier   night

'the illegal stalls were taken apart by the PP police unit last night'.
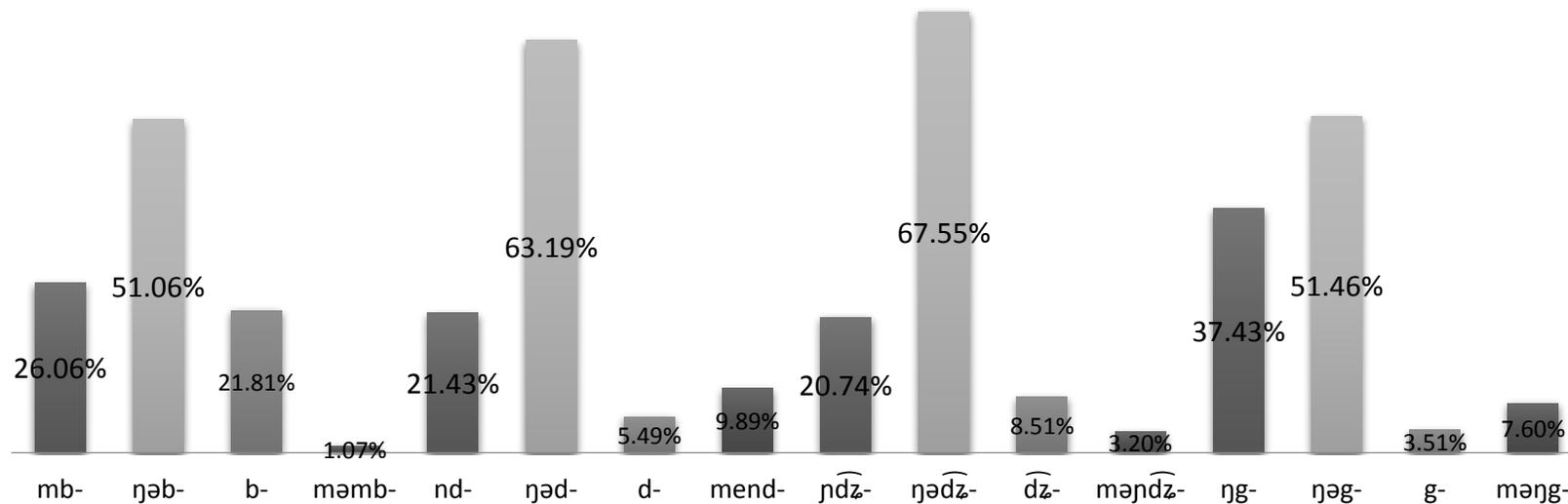
Expected Response:

Satpol PP **ŋəboŋəkarin/mboŋkarin** warung-warung liar itu tadi malem.

'PP police unite took apart the illegal stalls last night'

Production Task – inter-speaker variation
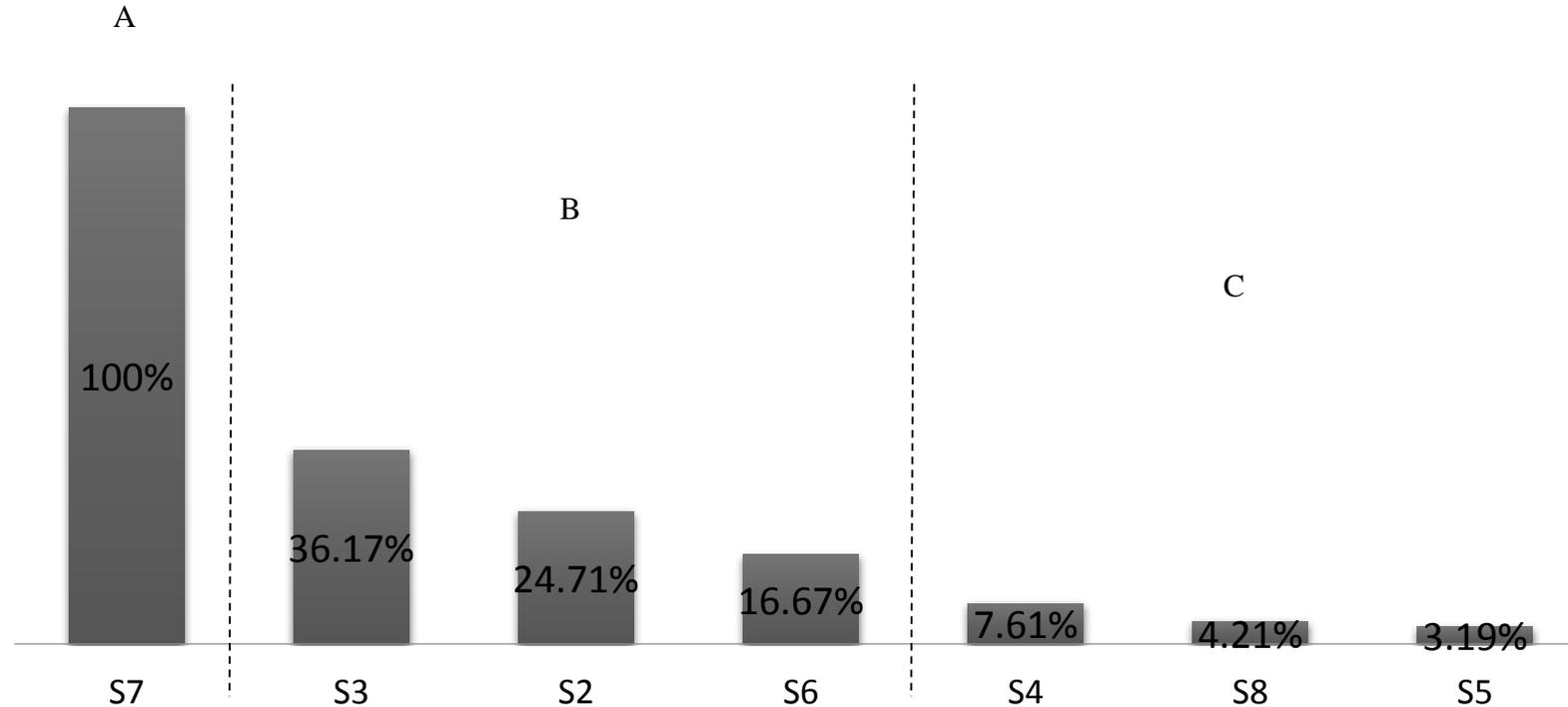
Production task - results

Production Task Result by place of articulation (pooled across 7 JI speakers)

- The epenthetic schwa forms are produced in greater numbers than the homorganic clusters, corresponding to the naturalistic data presented in the previous section

Production task - results

Results by 7 speakers (pooled across place of articulation)

- The figure presents the percentage of total homorganic clusters [mb-, nd-, ɲdʑ-] ,and [ŋg-] produced by each speaker.

Production task - results

- These results are consistent with the proposed hypothesis for this production task. In the hypothesis, it is predicted that <u>the epenthetic schwa forms have greater distribution than homorganic cluster forms.</u>

- These results also <u>corresponds to the results from JI corpus data</u> where epenthetic schwa has greater distribution than homorganic cluster forms.

- Here again, there are <u>two patterns observed</u> in this distribution.

  - They are non-variable speakers who consistently use schwa (C) and

  - Those with a significant pattern of variation (B).

- If additional speakers are added in the experiment, I would predict that more speakers of JI would show pattern C.

Production task – intra-speaker variation

Intra-speaker variation in production task

- The following table shows the intra-speaker variation that are produced by 7 speakers

Speakers' code	Gender	Variation within place of articulation	Percentage
S3	F	21	43.75%
S2	F	13	27.08%
S6	F	13	27.08%
S5	F	7	14.58%
S4	M	5	10.42%
S8	F	1	2.08%
S7	M	0	0%

Out of 48 test items (embedded in two differet sentence)

Summary from Production task

- Thus far, we have three sources of variation.:
  - Inter-speaker variation
  - Intra-speaker variation
  - Place of articulation.

VIII. Summary

Summary

- Based on the corpus results:
  - Betawi Malay and Jakarta Indonesian speakers produced two different patterns of variations of the N-prefix: inter- and intra-speaker variations.
    - For inter-speaker variation, BM speakers produced high numbers of the homorganic cluster forms [mb-, ɲd͡ʑ-] and [ŋg-] except for [nd-]. JI speakers produce higher numbers of the schwa epenthesic forms [ŋəb-, ŋəd-] and [ŋəg-] unless for [ŋəd͡ʑ-].
    - For intra-speaker variation, the corpus shows that the same individual speaker produced homorganic clusters and epenthetic schwa forms for the same lexical item.
- The results from production task that was conducted on JI speakers mirror the results from the corpus. Overall, as predicted in the hypothesis, JI speakers produced high numbers of epenthetic schwa forms than the homorganic cluster forms.
- Intra-speaker variation in both corpus and production task for JI demonstrate that the process of lexicalization has not occured. This is different from what Zuraw's (2010) finding based on her study in the written and dictionary corpus of Tagalog.

Summary Table

Source of variation	Varieties	Data
Lexical variation	Tagalog	Written corpus; dictionary; experiment
Inter-speaker variation	Jakarta Indonesian and Betawi Malay	Naturalistic spoken corpus
Intra-speaker variation	Jakarta Indonesian	Naturalistic spoken corpus; experiment
Place of articulation	Jakarta Indonesian	Naturalistic spoken corpus; experiment
	Betawi Malay	Naturalistic spoken corpus

Orientation for further research

- More production task data from BM and JI speakers are still needed to have better picture of the pattern of variation.

- It is also interesting to see whether or not gender of speakers could also be considered as another source of variation since Type B speakers consist only of female speakers.

- Robust distribution of [mb] by BM speakers also needs further account since this pattern could be a fact of current historical change that is imported from Javanese, another Austronesian language, that is closely related and geographically closed to BM community.

Reference

- Gil, D. and U. Tadmor. 2007. The MPI Betawi-Jakarta Database. A joint project of the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology and the Center for Language and Culture Studies, Atma Jaya Catholic University.

- Ikranagara, K. 1980. Melayu Betawi Grammar. *NUSA Linguistic Studies in Indonesian and Languages in Indonesia*, Vol. 9, Atma Jaya University, Jakarta.

- Lapoliwa, H. 1981. *A generative approach to the phonology of Bahasa Indonesia*. Canberra: Department of Linguistics, Research School of Pacific Studies, Australia National University.

- Muhadjir. 1981. Morphology of Jakarta Dialect: Affixation and Reduplication. *NUSA Linguistic Studies in Indonesian and Languages in Indonesia*, Vol. 11. Jakarta: Atma Jaya University.

- Pater, J. 2001. Austronesian nasal substitution revisited. In L. Lombardi, (ed.) Segmental phonology in Optimality Theory: Constraints and Representations. Cambridge University Press. 159-182.

- Pater, J. 1999. Austronesian Nasal Substitution and Other NC Effects. In Rene Kager, Harry van der Hulst, and Wim Zonneveld, eds., The Prosody Morphology Interface. Cambridge University Press. 310-343.

- Prince, A., and P. Smolensky (1993/2004) *Optimality Theory: Constraint interaction in generative grammar*. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published by Blackwell, 2004.

- Sneddon, J. 2006. *Colloquial Jakartan Indonesian*. Canberra, ACT: Pacific Linguistics.

- Zuraw, K. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28(2): 417-472.