Prosodic Phrasing in Mandarin Spontaneous Speech: A Realistic Account of a Clause-based Discourse Unit

Alvin Cheng-Hsien Chen
陳正賢

National Taiwan Normal University
alvinworks@gmail.com

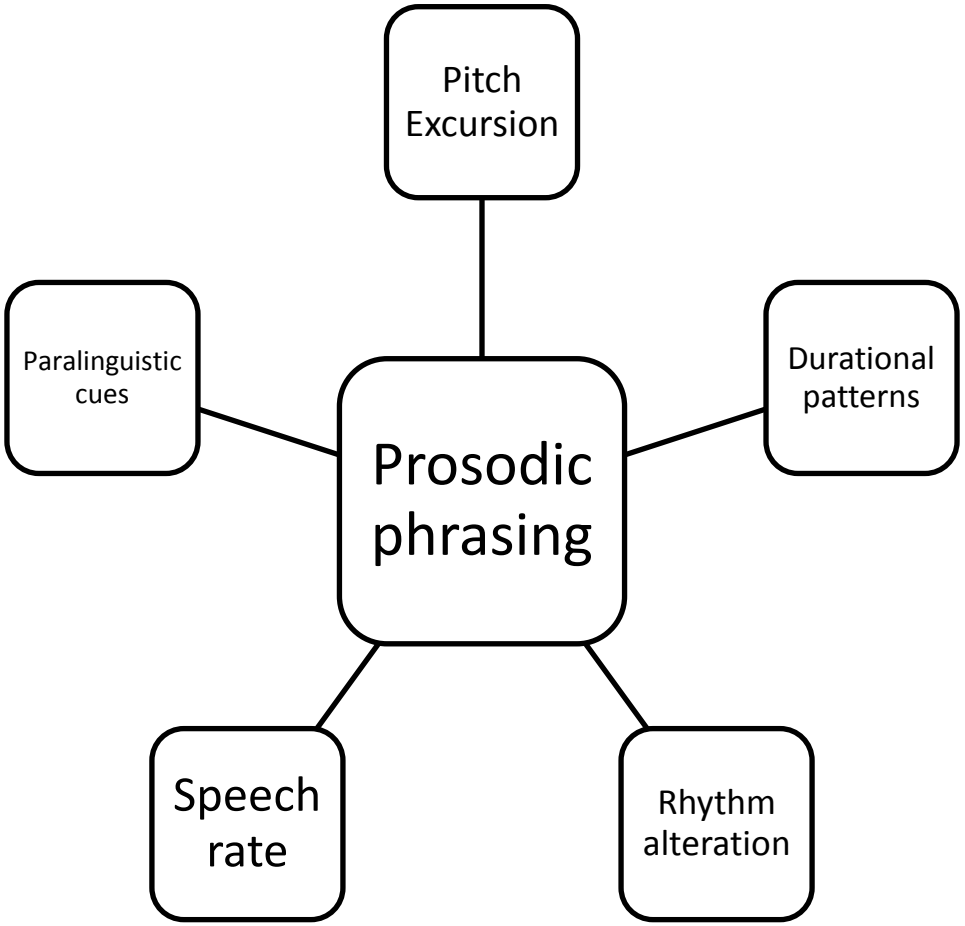
Department of Linguistics and Translations,
City University of Hong Kong, HK
09 Nov 2016

Background

General Understanding

- In speech production, we seem to have a basic production unit: prosodic units.
- In our grammar, we seem to have a basic grammatical unit: a clause-like unit.
- To find a basic unit in spontaneous speech production, one may start from one of these 2 perspectives without too much reliance on specific "theoretical frameworks".

A Prosody-Centric View



- Various prosodic cues
- Perceptually identifiable in speech production
- Less theory-dependent in locating the phrasing boundaries

A Brief Review on Prosody

- Discourse-oriented approach
  - Intonation-unit framework (Chafe, 1988, 1994; Croft, 1995; Iwasaki & Tao, 1993; Matsumoto, 2001; Ono & Thompson, 1996; Park, 2002; Tao, 1996; Thompson & Hopper, 2001)
  - The relation between IUs and various grammatical junctures (Phrases, Clauses, Paragraphs etc.)

- Phonology-based framework
  - Tone and Break Indices (ToBI) (Beckman & Hirschberg, 1994; Silverman et al., 1992)
  - Cross-linguistic adaptation for detailed prosodic transcription
  - Application in speech synthesis and other NLP tasks (Ostendorf & Veilleux, 1994; C.-y. Tseng et al., 2005; Wang & Hirschberg, 1992)

- Experimental approach
  - A study of the forms and functions of prosody
    - Prosodic forms: acoustic measures of durations, rhythms, pitch excursion, pauses etc.
    - Prosodic functions: syntactic boundaries, focus, prominence, contrastive stress
  - Consistency of the prosodic forms that speakers provide in conjunction with certain syntactic or pragmatic considerations

- Experimental approach
  - resolving local ambiguities in sentences (Kjelgaard & Speer, 1999; Warren, Grabe, & Nolan, 1995)
  - conjunction constructions (Clifton Jr. et al., 2006)
  - long-distance dependencies in complex sentences (Kraljic & Brennan, 2005; Schafer, Speer, Warren, & White, 2005; Snedeker & Trueswell, 2003)
  - focus and prominence in discourse (Wagner & Watson, 2010)
  - underlying syntactic structure (Fon et al., 2011; Steedman, 1991; C.-y. Tseng et al., 2005; Wagner, 2005)

Prosody-Grammar Alignment

- Where do speakers normally break in spontaneous speech?
- Typological studies (Croft 1995, Iwasaki and Tao 1993, Lin 2009, Matsumoto 2001, Park 2002 for Korean, Schuetze-Coburn 1994, Tao 1996)
- Prosody-Syntax Alignment
  - 55-60% of PUs are co-extensive with the *clause*
  - 40-45% of PUs mismatch with the *clause*

Complication

- Is 55-60% of the alignment between prosodic units and clause units enough as empirical evidence of the basic grammatical unit/schema, the "clause-based" unit?
- What about mismatches?
  - Internal syntactic configuration (Selkirk 1986)
  - Speech rhythm (Watson and Gibson 2004)
  - Interactional factors (Ono and Thompson 1995, Park 2002)
  - Performance arrangement (Ferreira 2007)

Objective

- To look for empirical evidence for grammatical constructions/schemas in the prosodic phrasing of speech production
  - Alignment or mapping between prosody and grammar
  - Grammatical configuration of the PUs
- To what extent the differing grammatical configurations of PU may contribute to systematic prosodic variation?

Outline

- Data
- Annotation
- Method
- Results
- Conclusion

Data

Annotation

Method

Results

Conclusion

Data

- Taiwan Mandarin Conversational Corpus (Tseng 2013)
  - Dr. Shu-Chuan Tseng at Academia Sinica
  - License Release: Sinica MCDC 8 (中研院漢語對話語音語料庫)
  - About 8 hours of conversation
  - 122k Words

Table 1. Corpus Description of the TMC Corpus.

Sub-Corpus	No. of Speakers	Length per conversation	Corpus Scenario	Conversation partners
MCDC	60 (37F, 23M)	1 hour	Free conversation	Strangers
MTCC	58 (33F, 25M)	20 minutes	Topic-oriented Conversation	Friends/ relatives
MMTC	52 (28F, 24M)	7 minutes	Map task dialogue	Friends/ relatives

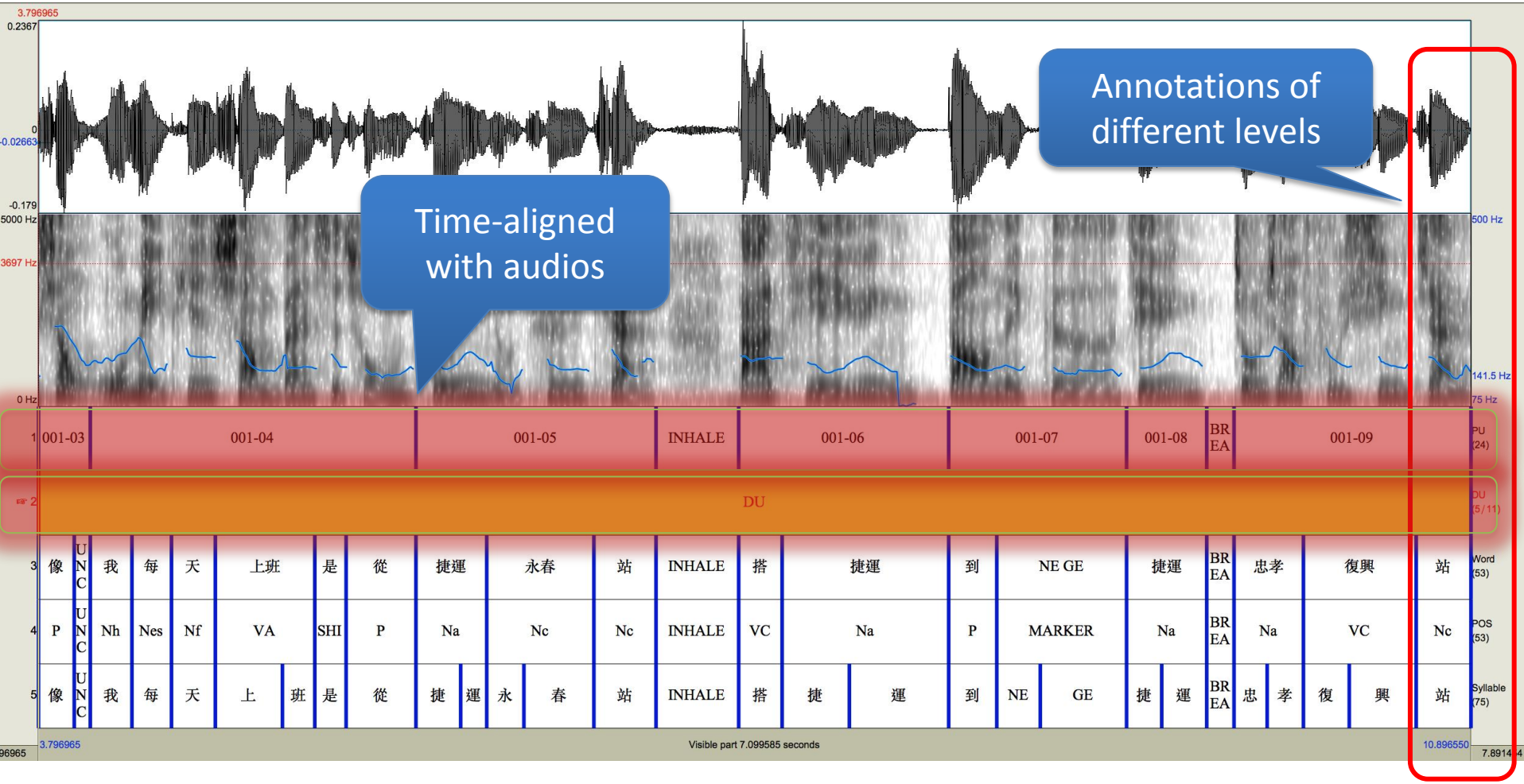
Sinica MCDC8 Subset

- Dataset for current study:
  - A subset of the Sinica MCDC 8 (中研院漢語對話語音語料庫)
  - 3.5 hours of face-to-face conversation
  - 16 Speakers
- Data size:
  - About 61k syllables
  - About 44k words
  - About 8500 Prosodic Units

Sinica MCDC 8 Annotations

- Prosodic Units
- Discourse Units
- Word
- POS
- Syllable/Phone

Data Format: Praat TextGrid



Data

Annotation

Method

Results

Conclusion

Prosodic Units

Prosodic Units

- In speech production, there seems to exist a kind of prosodic phrasing that is perceptually prominent cross-linguistically
- Terms
  - *tone unit* (Crystal, 1969)
  - *intonation group* (Cruttenden, 1997)
  - *intonation phrase* (Pierrehumbert, 1980)
  - *intonational phrase* (Nespor & Vogel 1986, Selkirk 1984)
  - *intermediate phrase* (Silverman et al. 1992)
  - *intonation unit* (Chafe 1994)

Intonation Units

- Wallace Chafe's IU
  - "a sequence of words combined under a single, coherent intonation contour" (1987:22)
- Features for identifying boundaries between IUs (Chafe 1994: 58)
  - Changes in fundamental frequency, or pitch
  - Changes in duration or tempo (manifesting itself as shortening and lengthening of syllables or words)
  - Changes in intensity or loudness (including stress and accents)
  - Alterations between vocalization and silence(pausing)
  - Changes in voice quality (creaky voice
  - Changes in speaker turn

Prosodic Units

- We adopt "Prosodic Unit" because of its more general construct for including other prosodic patterning in addition to intonation.
- Operational criteria (Liu and Tseng 2009)：
  - Pitch reset
  - Lengthening
  - Occurrences of paralinguistic cues
  - Alteration of speech rate

Inter-labeler Agreement

Liu and Tseng 2009:

- 3 Annotators
- 150 speaker *turns*
- Each annotator's result is compared to the *finalized* annotations for Precision and Recall.

	Labeler-01	Labeler-02	Labeler-03
# of PUs labeled	210	217	213
# of finalized PUs	218	218	218
# of correctly labeled PU-final boundary compared with finalized PUs	196	207	195
Precision rate%	93%	95%	92%

Table 1-1: Precision rate of the prosodic segmentation (Table 1 in Liu and Tseng 2009)

	Labeler-01	Labeler-02	Labeler-03
# of PUs labeled	210	217	213
# of consistent PU-final boundary	178	178	178
Consistent rate %	85%	82%	84%

Table 1-2: Inter-labeler's consistency (Table 2 in Liu and Tseng 2009)

- **Precision** (The percentage of how many PUs (in the final set) were labeled by Labeler X) → around **90%**
- **Recall** (The percentage of how many PUs (in the final set) were labeled by ALL Labelers) → around **82%**

Practical Values of PUs

- Better segmentation units in NLP
  - In an automatic POS tagging experiment, it is demonstrated that transcripts with annotations of prosodic boundaries achieved a slightly better performance than the original transcripts with only the speaker turn annotation. (Liu and Tseng 2009)
- Tailored to spontaneous speech processing
  - Disfluencies
  - Hesitations
  - Repairs

Discourse Units

Discourse Unit

- The objective is to look for a basic grammatical unit in spontaneous speech
- "Basic unit" in SS is less operationally defined across different studies.
- A notional equivalent of the "clause," more defined in written grammar, is often a practical start.

A Common Solution

- A proposition-based unit works well in many discourse-based studies (Croft, 1995; Givón, 1984; Halliday, 1989; Huang & Chui, 1997; Langacker, 2001; Lehmann, 1988; Matsumoto, 2000; Park, 2002; Tao, 1996; Thompson & Couper- Kuhlen, 2005; Thompson & Hopper, 2001)
- A Socio-cognitive basis for "proposition-based units" in discourse
  - The most frequently use "format" to perform social actions (Thompson & Couper- Kuhlen, 2005)
  - A primitive unit to express one event (state of affair)

Operational Criteria

- A Discourse Unit (DU) is a unit where
  - speakers talk about some entity, often via the ***Subject*** (e.g. people, things, events, states, abstraction) as their starting point and,
  - add information about that entity via the ***Predicate***.
- It is due to this nature of single predication that a DU has become "the locus of the densest network of distributional and dependency relationship" in most syntactic theorizing (Miller & Weinert 1998:77)

Operational Criteria

- Decision of the “main predicate”
  - To ensure reliability and consistency of our annotation
  - Chinese PropBank Framesets
    - Frames of the main predicates
    - Propositional structure
    - Projected boundaries for DUs
- A “clause-based” Discourse Unit:
  - Accommodation for the nature of spontaneous speech (Prevot et al 2015)

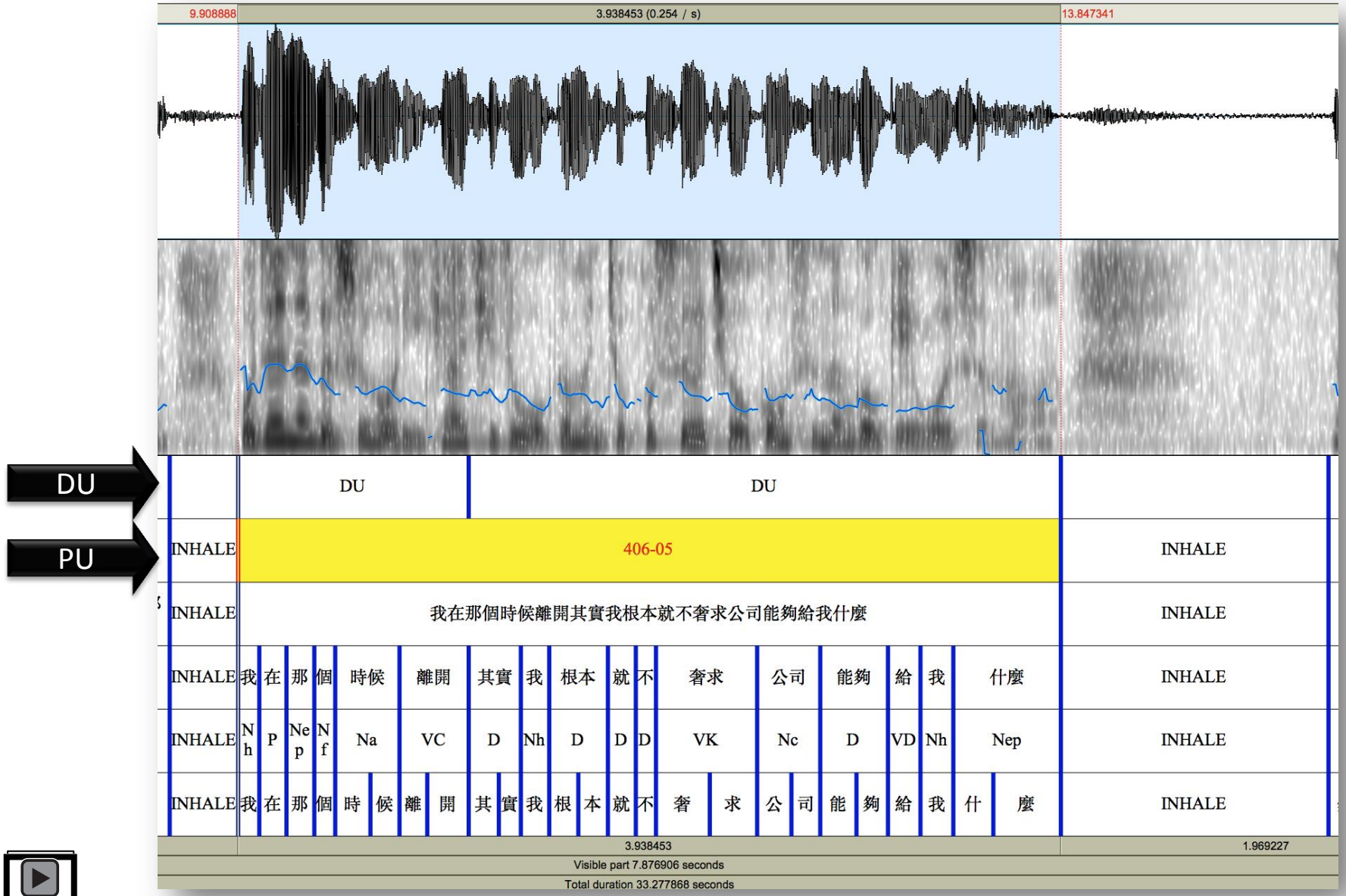
Issues in DU Segmentation

- Verbal Complex
  - Complement-taking verbs
    - Modality verbs
    - Manipulative verbs
    - Perception-cognition-utterance verbs
  - Serial Verb Construction (Baker 1989, Givón 1991)
- Grammaticalization
- Language-specific constructions
- Unique patterns in spontaneous speech

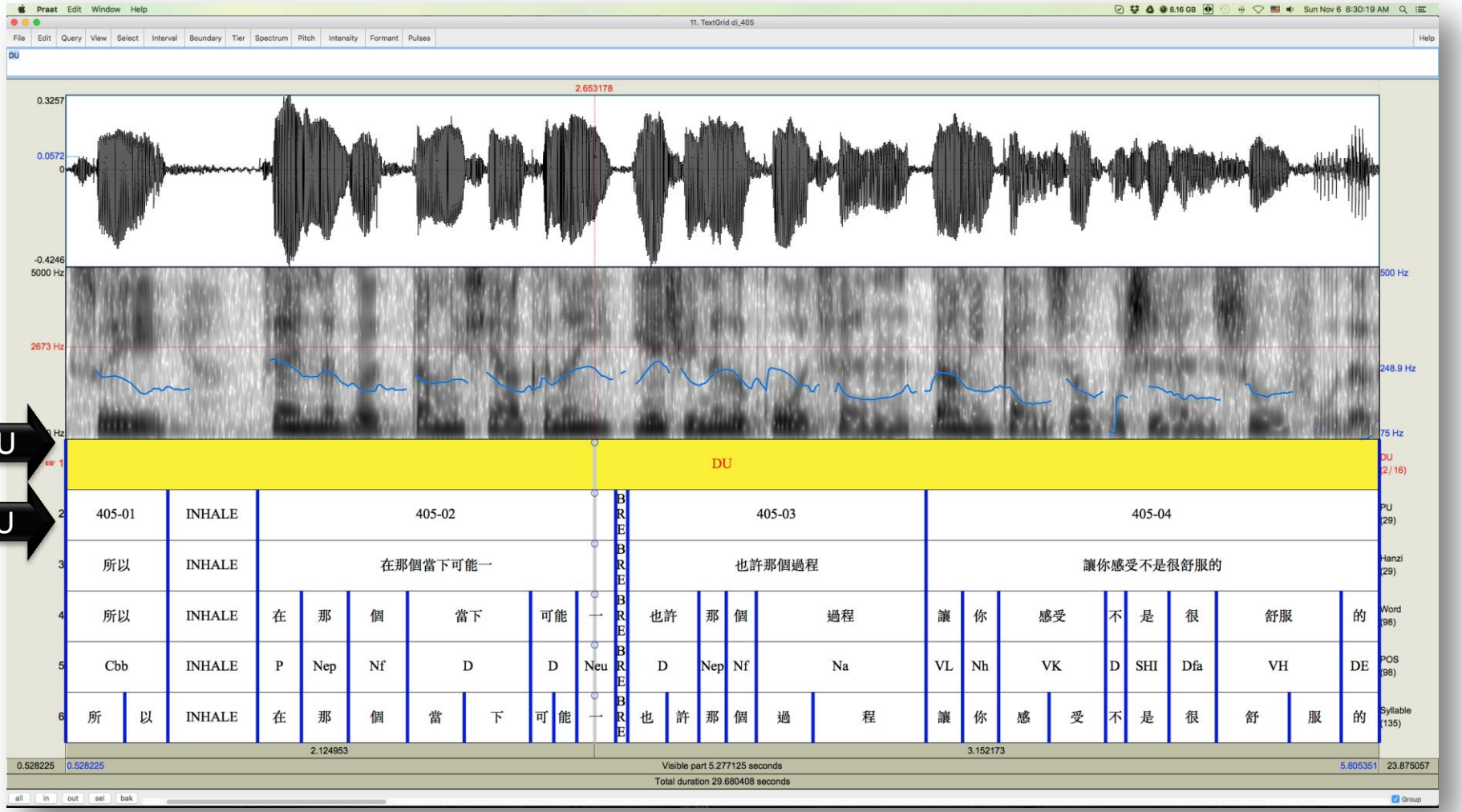
Inter-labeler Agreement

- Kappa Coefficient: 0.86 (Prevot et al. 2015)
  - 2 Labelers
  - About 20% of the dataset were annotated by 2 labelers for an annotation agreement test
  - For each word boundary, we ran the agreement test using Kappa coefficient for the binary labels (DU vs. non-DU boundaries)

Example: PU across DUs



Example: DU across PU



Data

Annotation

Method

Results

Conclusion

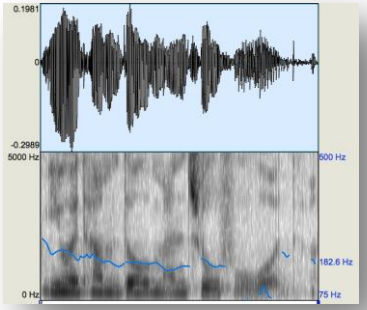
Objective

- Given a clause-based DU schema, how its interaction with PUs may contribute to a systematic variation in the prosodic structure of PU?
- If there is a strong correlation, this may serve as empirical evidence for how a grammatical schema emerges as a realistic unit in speech production.

Grammatical Configuration → Variation in Acoustic Measures

Acoustic Measures

Acoustic Representation



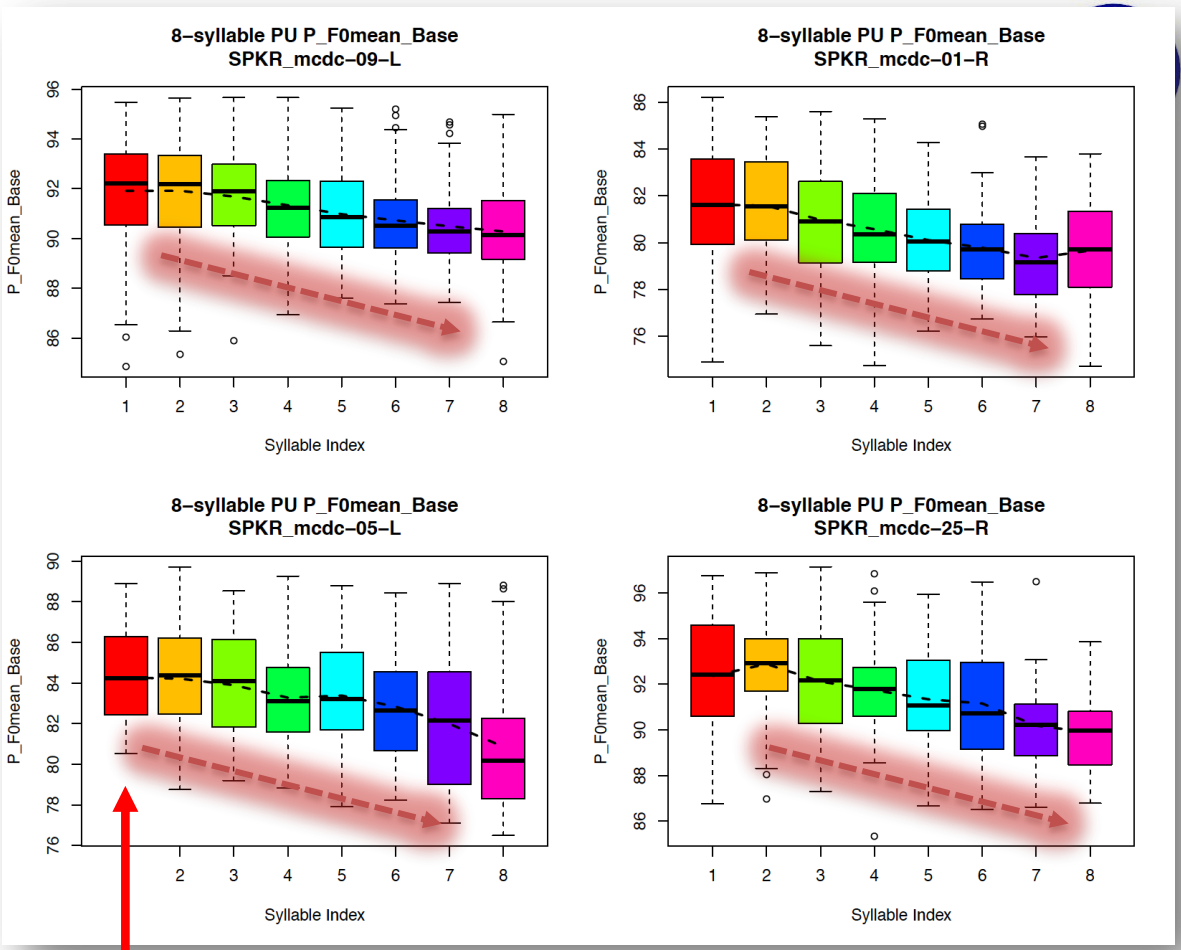
- Pitch Variation
  - Initial Pitch
  - Final Pitch
- Duration Pattern
  - Initial Normalized Duration
  - Final Normalized Duration
- Rhythmic change
  - Initial PVI
  - Final PVI
- Speech Rate
  - Speech Rate
- Paralinguistic
  - Pauses

Pitch Variation in PU

- Criteria for PU annotation
  - Changes in fundamental frequency, or pitch
  - Pitch reset
- A general tendency
  - Pitch is typically raised in the discourse initial position and lowered in the discourse final position (Shih 2000)

F0 Declination in PU

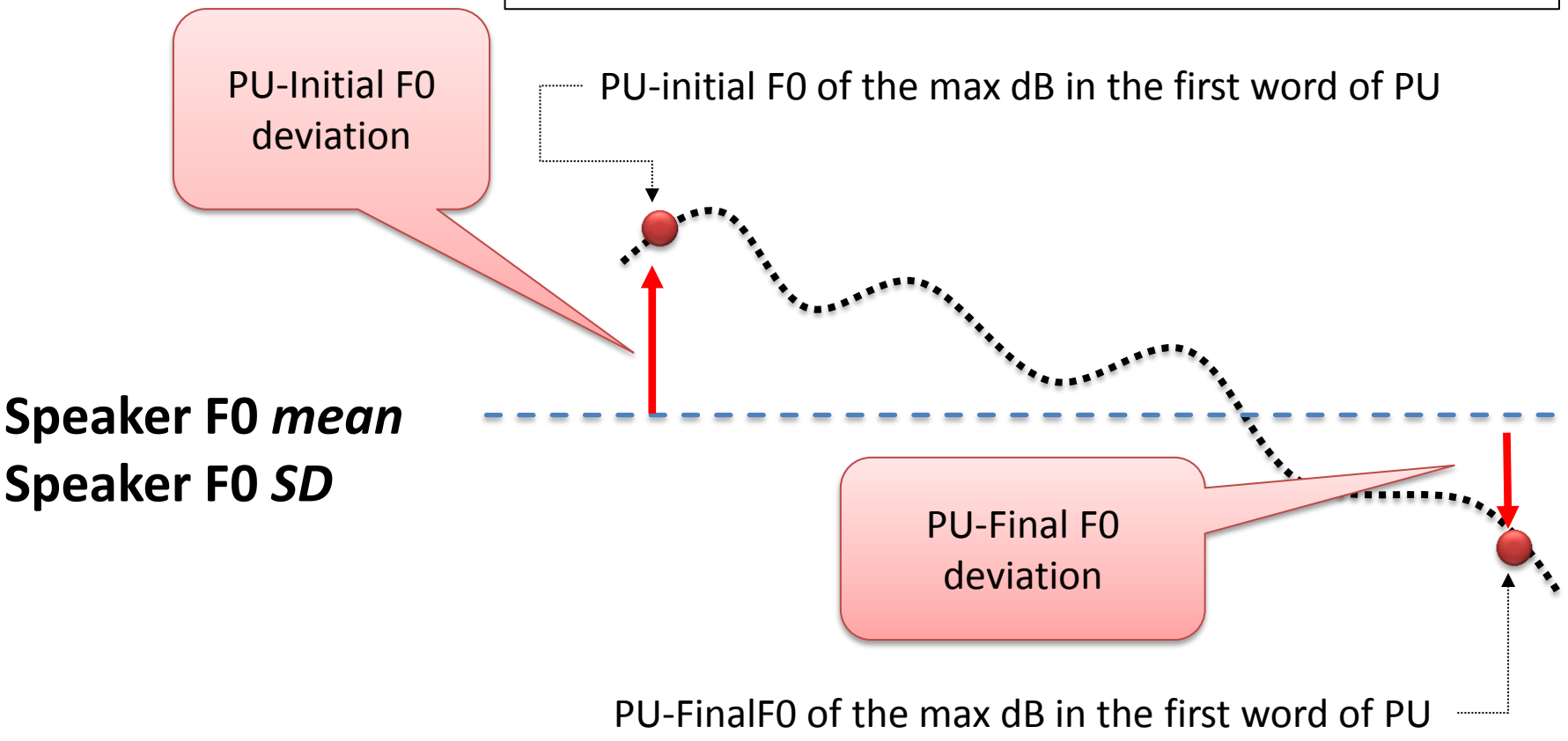
- 4 speakers in our data
- For all their 8-syllable PUs
- For each i^{-th} syllable, we plot the distribution of the F0 means (i.e. Boxplot).
- Downward F0 movement is prominent.



This would represent the F0 mean distribution for all the FIRST syllables of the 8-syllable PUs produced by SP05

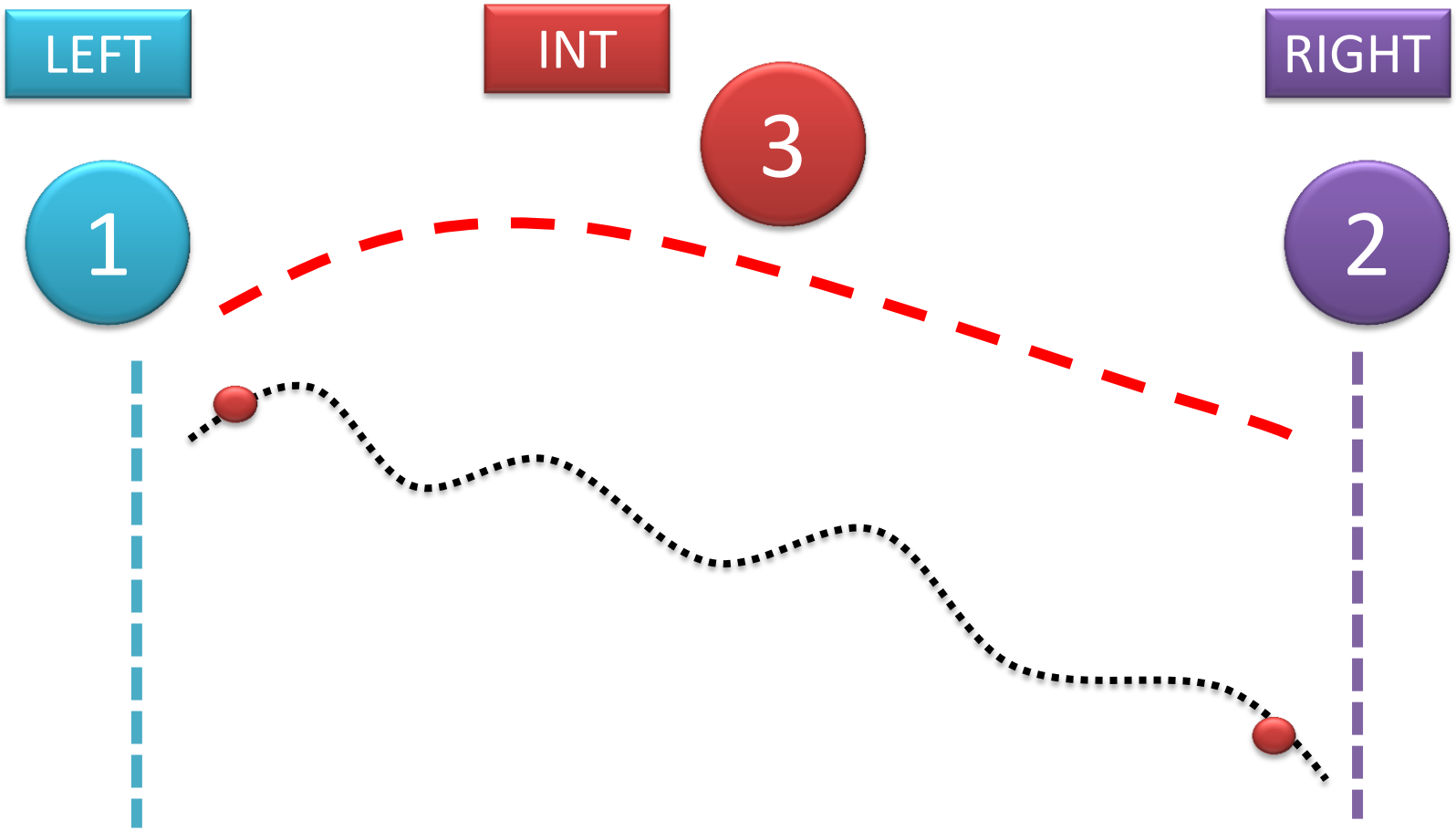
Acoustic Measure

- For each prosodic unit
  - PU-Initial F0 Deviation (**Initial F0**)
  - PU-Final F0 Deviation (**Final F0**)



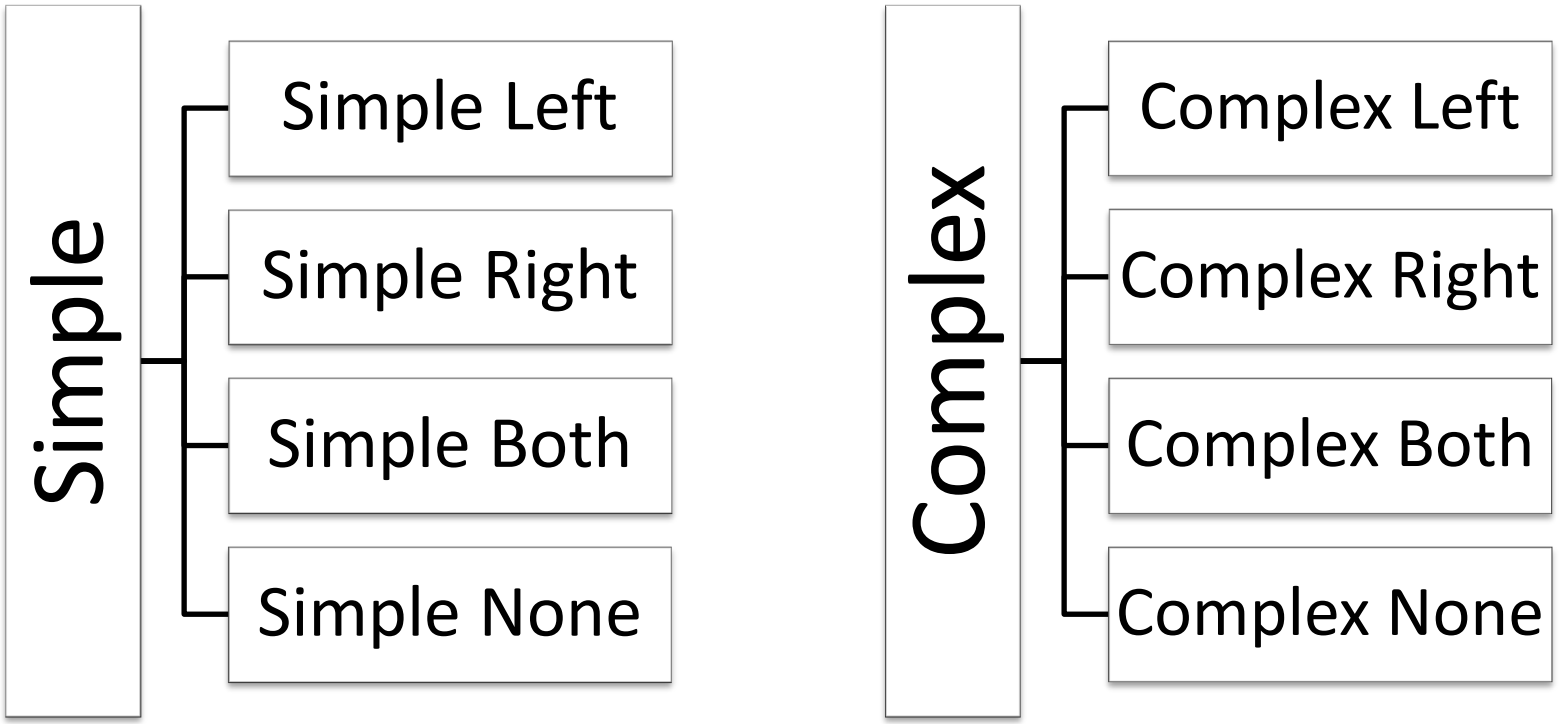
Grammatical Configuration

Grammatical Configuration (DU-PU)



Grammatical Configuration of PU

INT x LEFT x RIGHT



Research Question

- Is there a correlation between **PU-initial F0** and PU grammatical configuration in terms of **LEFT, RIGHT, INT**?
- Is there a correlation between **PU-final F0** and PU grammatical configuration in terms of **LEFT, RIGHT, INT**?

Hypothesis

- In general, PU exhibits a prosodic pattern of **initial F0 higher** than the baseline; the **LEFT** may strengthen this tendency.
- In general, PU exhibits a prosodic pattern of **final F0 lower** than the baseline; the **RIGHT** may strengthen this tendency.
- There is a correlation between the F0 variation and the **LEFT, RIGHT, INT** and their **Interactions**.

Linear Mixed Effect Model

Prosodic Structure

- PU-Initial F0
- PU-final F0

~

Grammatical Configuration

- Fixed Effects:
  - **LEFT**
  - **RIGHT**
  - **INT**
  - **LEFT:RIGHT**
  - **LEFT:INTDU**
  - **RIGHT:INTDU**
- Random effects:
  - Subjects (18 SPs)
  - PU Length (Num of W)

Data

Annotation

Method

Results

Conclusion

Descriptive Statistics

Descriptive Statistics: INT

INT	N	%	Average of wordnum
Simple	7430	86.77%	3.80
Complex	1133	13.23%	8.67
Total	8563	100.00%	

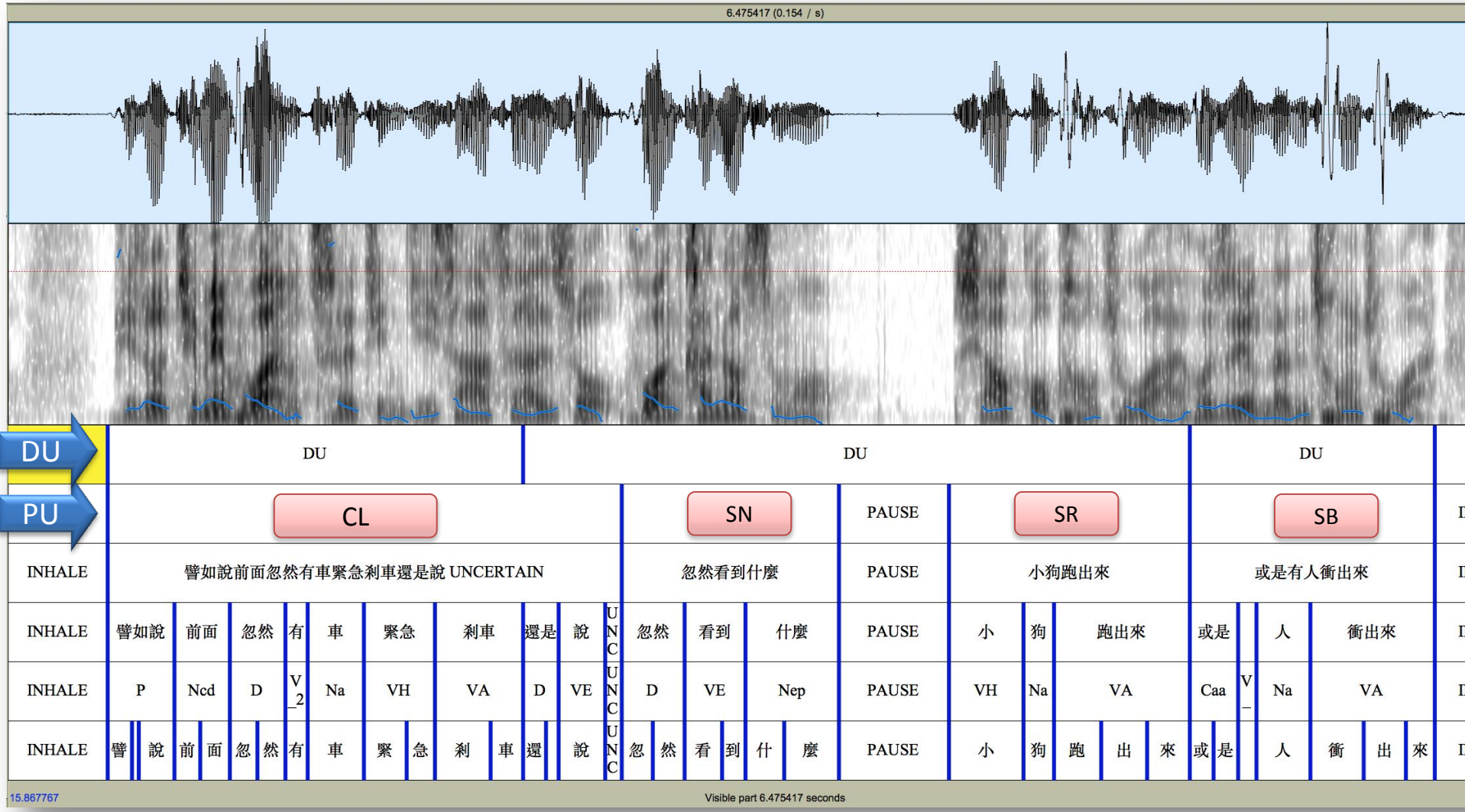
- A great majority of PUs are Simple PUs (INT = 0) (DU sub-components)
- About 13% of the PUs are Complex PUs, integrating more than one DU.
- As expected, Complex PUs are about twice the length of the Simple PUs

Descriptive Statistics: LEFT x RIGHT

	RIGHT					
	0		1		Total N	Total %
LEFT	N	%	N	%		
0	1731	20.21%	2062	24.08%	3793	44.30%
1	2065	24.12%	2705	31.59%	4770	55.70%
Grand Total	3796	44.33%	4767	55.67%	8563	100.00%

- About 56% of the PUs finish at the DU boundaries.
- About one-third of the PUs are fully co-extensive with the DUs.
- About 80% of the PUs are aligned with the DU boundaries on at least one end.
- Such a tendency exists across Simple and Complex PUs.

Example of LEFT x RIGHT x INT



PU-Initial

PU-Final

Statistical Results

PU-Initial F0 Deviation

Hypothesis (Recap)

- In general, PU exhibits a prosodic pattern of **initial F0 higher** than the baseline; the **LEFT** may strengthen this tendency.
- In general, PU exhibits a prosodic pattern of **final F0 lower** than the baseline; the **RIGHT** may strengthen this tendency.
- There is a correlation between the F0 variation and the **LEFT, RIGHT, INT,** and their **Interactions**.

Initial F0

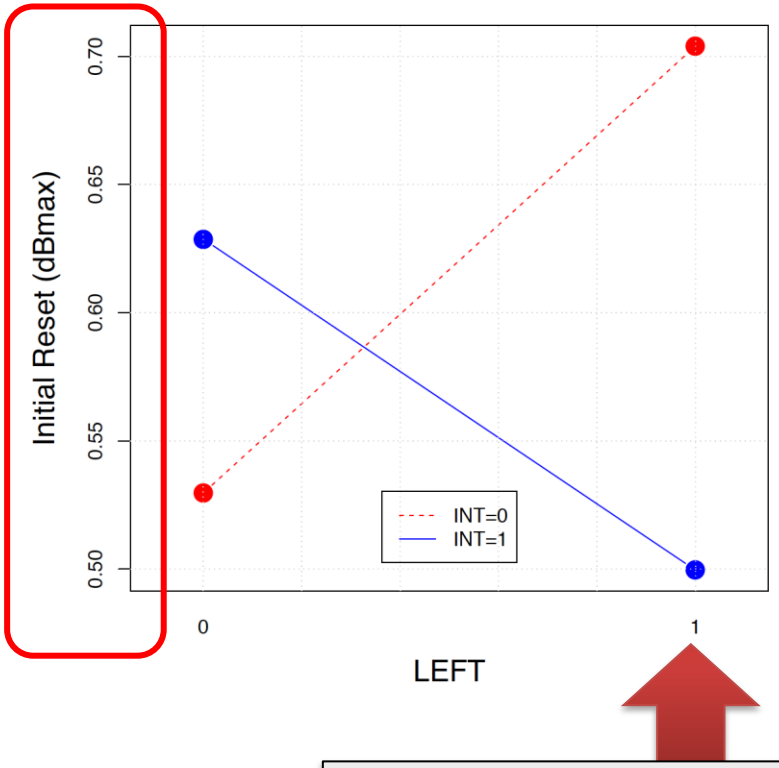
- 2 Interaction Effects on PU-initial F0:
  - LEFT*INT (β = -0. 3015, $p < 0.01$)
  - RIGHT*INT (β = 0.3261, $p < 0.01$)
- General Tendency
  - DU boundary effects (LEFT and RIGHT) on the prosodic structure (Initial F0) may differ for Simple and Complex PUs.

Interaction (1): LEFT x INT

Initial F0: LEFT*INT

- LEFT has different effects on **Simple** and **Complex** PUs in terms of the PU-initial F0 deviation
- LEFT has a strong effect on the *increase* of the **Simple** PU-initial F0, inflating the expected initial F0 deviation.
- For **Complex** PUs, non-LEFT introduces more initial F0 deviation.

The positive F0 means suggest that in general the initial F0 is above the baseline



We expect LEFT would strengthen this tendency

Complex non-LEFT PUs

- Discourse Conjunctions

  [#di_003]

  – <DU> NA 如果

  – 從 南港 過去 <DU> 要 怎麼 去 <DU>

- Planning Process

  [#di_017]

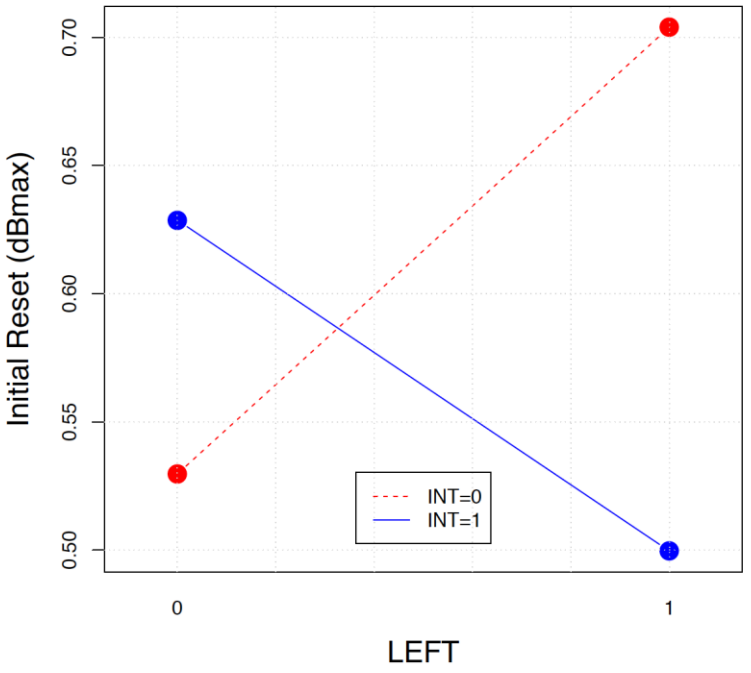
  – <DU> 在 橋 上面 會 有 一點

  – 塞 LA <DU> 不過

  – 上 了 橋 以後 <DU> 就 蠻 順 的 <DU>

Initial F0: LEFT*INT

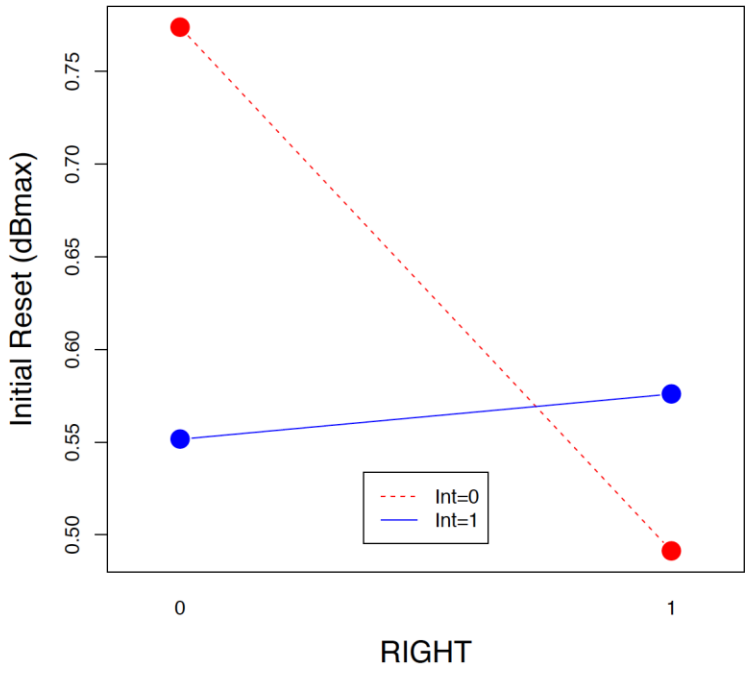
- For a **Complex** PU that is *not left-aligned*, the preceding PU often serves as a buffer for complex events structuring (e.g. hesitation, conjunctions, disfluencies)
- The higher F0 in non-left-aligned Complex PU may suggest a ready-state for the up-coming of the complex events.



Interaction (2): RIGHT x INT

Initial F0
RIGHT*INT (1)

- Initial F0 deviation is NOT often discussed in terms of its correlation to the PU-final alignment in literature.
- RIGHT has a strong effect on **Simple** PUs that the right alignment reduces the scale of Initial F0 deviation.
- PU-Initial F0 deviation correlates with whether a **Simple** PU is going to end a proposition.



PU-Final F0 Deviation

Hypothesis (Recap)

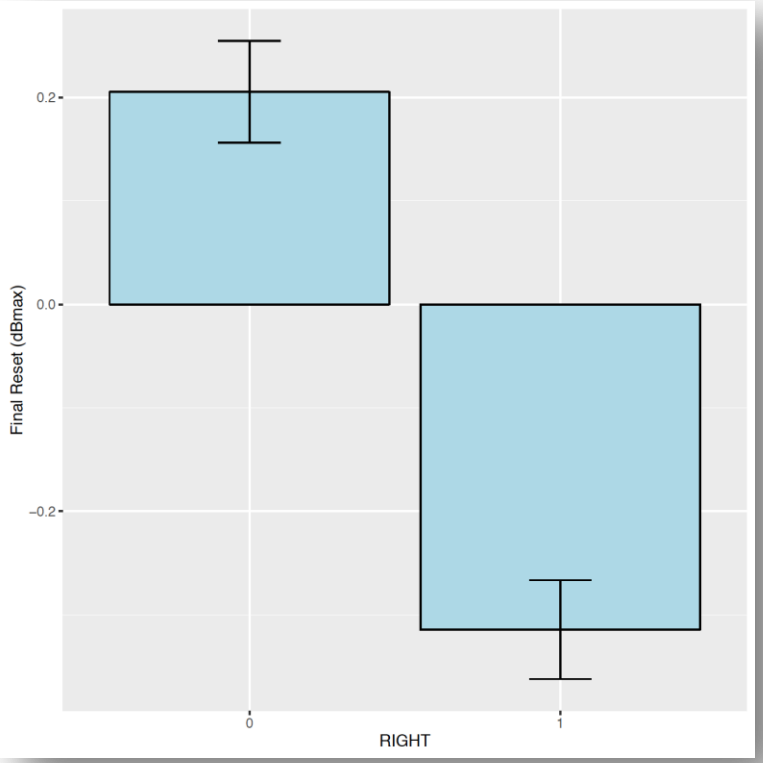
- In general, PU exhibits a prosodic pattern of **initial F0 higher** than the baseline; the **LEFT** may strengthen this tendency.
- In general, PU exhibits a prosodic pattern of **final F0 lower** than the baseline; the **RIGHT** may strengthen this tendency.
- There is a correlation between the F0 variation and the **LEFT, RIGHT, INT**, and **their Interactions**.

Final F0

- 2 Main Effects on PU-final F0
  - LEFT ($\beta = 0.1468$, $p < 0.01$)
  - RIGHT ($\beta = -0.3605$, $p < 0.01$)
- Highlights
  - RIGHT is as expected inflating the lowering effect of the PU-final F0 deviation.
  - LEFT is more interesting.

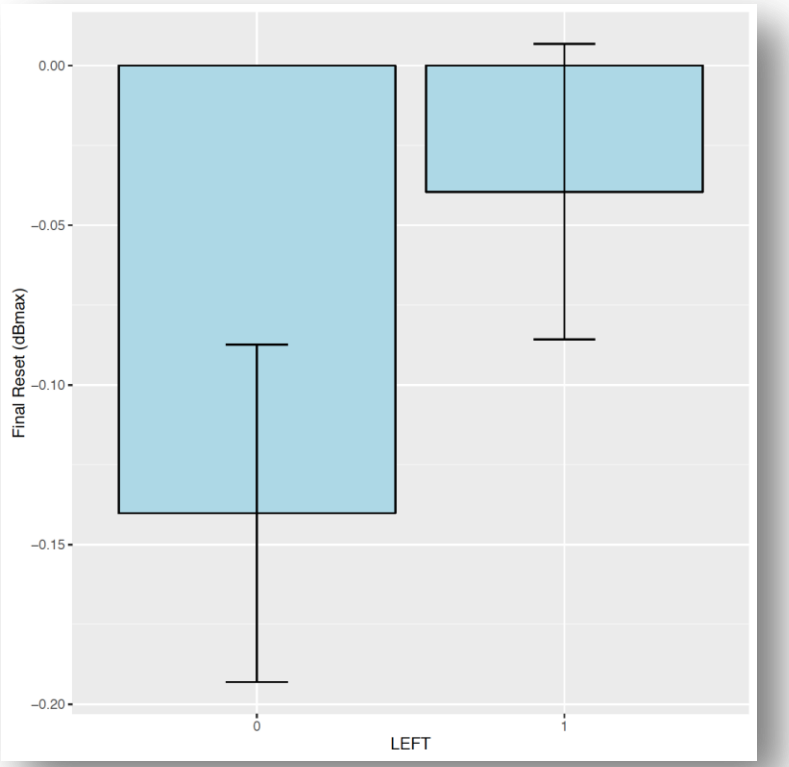
Final F0: RIGHT

- Bar Plot for F0 Means
  - Bars represent the F0 means for each level of RIGHT.
  - Whiskers = CI of the means
- It's obvious that if a PU ends at a DU boundary, the Final F0 is much lower than one's baseline.



Final F0:
LEFT

- In general, final F0 tends to be lower than SP baseline
- When LEFT = 0, the final F0 is even much lower than SP baseline
- When LEFT = 0, it is more likely to be a DU-internal PU, thus being in the later stage of the discourse structure.
- When LEFT = 1, it is the DU-initial PU, thus at the beginning of discourse structure.



Data

Annotation

Method

Results

Conclusion

Findings Summary

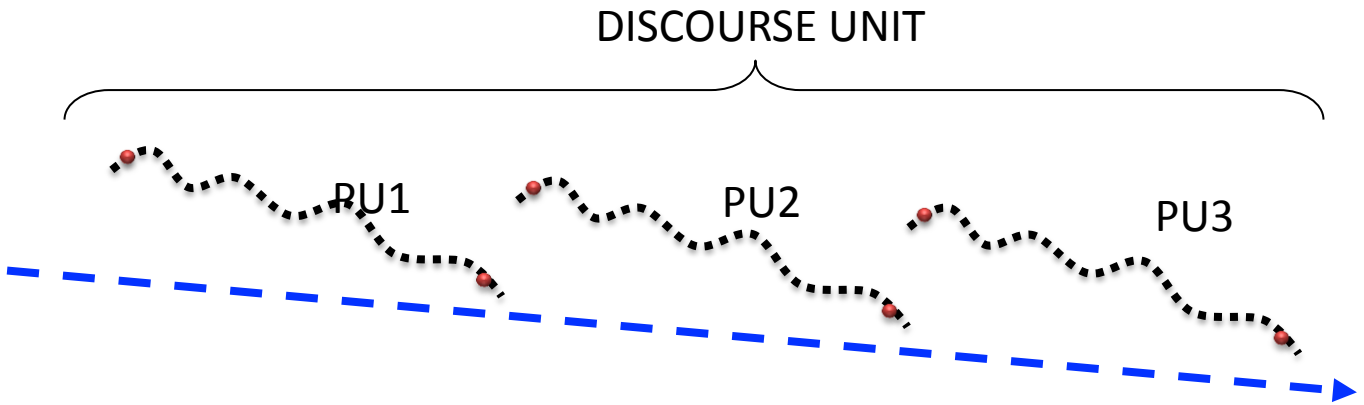
- **Initial F0**
  1) LEFT strengthening effect on PU-initial F0 only correlates with Simple PUs (cf. LEFT*INT)
  2) RIGHT also correlates with initial F0 in that for simple PUs whether PUs are going to end at a DU boundary is anticipated in the Initial F0. (cf. RIGHT*INT)
- **Final F0**
  1) RIGHT indeed shows a strong correlation with a stronger lowering effect on PU-final F0.
  2) LEFT also correlates with final F0 in that the relative position of a PU in a discourse structure is reflected in the Final F0.

Implication (1)

- The Initial F0 correlates with **RIGHT**
  - At the onset of the PU, SP has already planned a primitive sketch of *the intended DU*, whose **completeness** is anticipated in the degrees of PU-initial F0 deviation.
- The Final F0 correlates with **LEFT**
  - At the end of the PU, SP finishes the PU with the **previous knowledge of the primitive sketch** of *the intended DU*, which is reflected in the degrees of PU-final F0 deviation.

Implication (2): F0 Declination

- If a PU is not left-aligned, it is a DU-internal PU.
- The correlation between LEFT and PU-final F0 may serve as indirect evidence for a general trend of F0 declination in discourse structure.
- The later the position of the PU in the discourse unit, the more the PU-Final F0 deviates from the baseline.



Incremental Speech Production?

- When a speaker is formulating the morpho-phonological encoding and articulating, they are capable of conceptually planning the upcoming words at the same time.
- This "look-ahead" conceptual planning in articulation may be supported by the acoustic measures of PU.
  - Initial F0 <-> RIGHT
  - Final F0 <-> LEFT

Acknowledgement

- Dr. Shu-Chuan Tseng and her research team in Academia Sinica
- Research Grants of the Ministry of Science and Technology Taiwan

Q & A
Thank you

Alvin Cheng-Hsien Chen
陳正賢

National Taiwan Normal University
alvinworks@gmail.com

09 Nov, Department of Linguistics and Translations,
City University of Hong Kong, HK