Use of Classifiers in Language Mixing: The case of the Cantonese–Japanese Variety

The Workshop on Altaic Formal Linguistics

Minnie Pui Yee Yuen

Humboldt University of Berlin
Leibniz-Zentrum Allgemeine Sprachwissenschaft

30th June – 2nd July 2026



Leibniz-Zentrum
Allgemeine Sprachwissenschaft

Outline

1 Background information
- Theoretical Framework
- Classifiers

2 The *n*-approach
- Classifier Information at *n*
- The Current Study

3 Study
- Elicitation Implementation
- Result & Discussion

4 Implications

5 References

Theoretical Framework: Distributed Morphology

- Distributed Morphology (Halle & Marantz, 1993; Embick, 2015):
  - Decompositional and late insertion model
  - All linguistic elements are syntactically generated, including words
  - An acategorial $\sqrt{Root}$ (free from all grammatical information) + a categoriser (e.g. *n*, *v*, *a*)
  - Three lists:
    1. List 1: $\sqrt{Root}$ and Synsem features, input of syntactic operation
    2. List 2: Vocabulary Insertion, phonological exponents
    3. List 3: Encyclopedia

The $\sqrt{Root}$ and the *n*

$\underline{\sqrt{Root}}$

- Category-less
- Free from all syntactic information, e.g. gender
- A conceptual core (void of all semantic information) (Acquaviva, 2009; Harley, 2014; Arad, 2003)[1]

$\underline{n}$

- Word-maker + category: turn a $\sqrt{Root}$ into a noun
- Functional layer: with syntactic information on it, e.g. grammatical gender (Kramer, 2015; Embick, 2015)
- But *how about the semantic selectional information related to classifiers?*

[1] Marantz (2001) proposes otherwise

Previous Study on Language Mixing

- Mixing between *n* and $\sqrt{Root}$ is possible and frequent
  - English–Norwegian (Alexiadou & Lohndal, 2018; Grimstad et al., 2015; Riksem, 2018)
  - Greek–German (Alexiadou & Lohndal, 2018)
  - Greek–English (Alexiadou & Lohndal, 2018)
  - Spanish–German (López, 2020)
  - Spanish–English (López, 2020)
- López (2020): only one computational system, hence one List I and one List II for bilinguals, and $\sqrt{Roots}$ shared between languages

Classifier

- Cantonese (YUE) and Japanese (JP) are classifier (CL) languages
  - CL is (almost) obligatory for nouns (except: generic reading)
  - CL has multiple syntactic functions: classification, counting, pluralisation, specificity, definiteness

Classification I

Classification:

- Each noun gets at least one CL
- holds a semantic relationship (usually) according to the inherent characteristics, e.g. shape, size, function, etc.
  1. Rigid (the noun only has one CL)
  2. Flexible (the noun can take more than one)
  3. Arbitrary (the relationship is opaque)
  4. Default (no semantic relationship)

Classification II

- **Rigid**: a restricted selectional relationship

(1) a. 一 本/ * 張/ * 架/ * 隻/ * 枝 書
jat^{55} *bun*^{35}/ *zoeng*^{55}/ *gaa*^{33}/ *zek*^{33}/ *zi*^{55} *syu*^{55}
one CL.[*book*]/ CL.[*piece*]/ CL.[*device*]/ CL.[*animal*]/ CL.[*long*,*fix*] book
'one book' [Cantonese]

b. 本 一 冊/ * 本/ * 枚/ * 頭/ * 匹
hon ichi satsu/ hon/ mai/ tou/ hiki
book one CL.[*book*]/ CL.[*prolonged*]/ CL.[*flat*]/ CL.[*big.animal*]/ CL.[*small.animal*]
'one book' [Japanese]

Classification III

- **Flexible**: the choice of CL is relatively free, normally depends on its "highlighted" feature, e.g. shape

(2) a.

一	張/	捲/	疊/	條	紙
jat^{55}	*zoeng*^{55}/	*gyun*^{35}/	*dap*^{22}/	*tiu*^{23}	*zi*^{35}
one	CL.$_{[flat]}$/	CL.$_{[roll]}$/	CL.$_{[stack]}$/	CL.$_{[prolonged]}$	paper

Reading: 'one piece/ roll/ stack/ long piece of paper' [Cantonese]

b.

紙	一	枚/	巻/	束
kami	ichi	mai/	maki/	taba
paper	one	CL.$_{[flat]}$/	CL.$_{[roll]}$/	CL.$_{[bundle]}$

Reading: 'one piece/ roll/ stack of paper' [Japanese]

Classification IV

- **Arbitrary**: no transparent reason of the CL choice

(3) a. 一 把 口
jat^{55} *baa*^{35} *hau*^{35}
one CL.$_{[handle]}$ mouth
'a mouth' [Cantonese]

b. 一 本 の 歯
ichi hon no ha
one CL.$_{[prolonged]}$ GEN tooth
'One tooth' [Japanese]

- *baa* is the CL for objects with a handle (e.g. *a knife*, *a sword* etc.). But *mouth* combines with *baa*, when a *mouth* does not have a handle
- *hon* is the CL for long and skinny objects (e.g. *a bottle of beer*, *an umbrella*) and *tooth* does not seem to associate with this feature

Classifier Selectional Information and *n*

- If $\sqrt{Root}$ is free from syntactic and semantic information, the semantic selectional information for classifiers cannot come from $\sqrt{Root}$
- Instead, the information is from the *n*, just like grammatical gender, proposed by Kramer (2015)

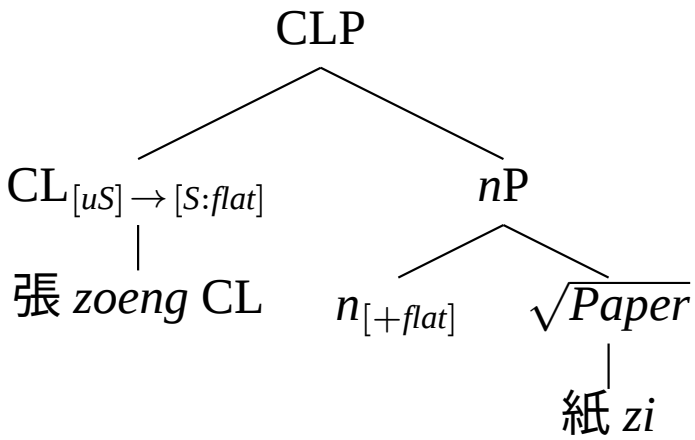
Parallel with gender information

- Similarities between gender assignment and the classifier selection information (Aikhenvald & Mihas, 2019; Aikhenvald, 2025):
  1. **Exclusive** for the nominal domain: gender assignment and selectional information are exclusive to the nominal domain
  2. For **classification**: for classifying nouns into groups
  3. Being **transparent and arbitrary**: either transparent (natural, predictable) or arbitrary (random)
  4. **Active in syntax and semantics**: active in agreement with other elements in syntax and semantically active for interpretation

Syntactic Representation

- Selectional information: represented as a feature, e.g. $n_{[+flat]}$
- The classifier phrase:
  - an undefined feature (probe)
  - seeks value in the c-command domain
  - Agrees with the value at the *n* (goal)

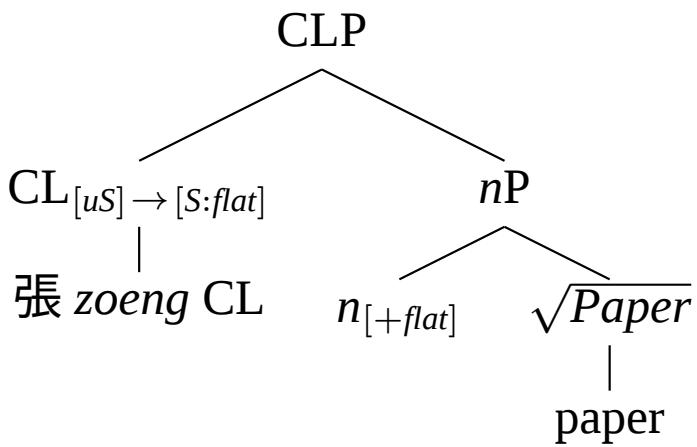
(4)



Previous work

I have worked with a language mixing variety involving one classifier language and one non-classifier language: Cantonese–English, the result aligns with the *n*-approach – English nouns can take a Cantonese-specific classifier, even though it is not a classifier language

(5)



The Current Study

In language mixing between two classifier languages, the shared List I consists of all the *n*s from both languages, indicating that the two classifier systems can interact.

Design I

(i)Same feature–Same CL	(ii)Same feature–Different CLs	**(iii)Different features–Different CLs**
professor 名 [*human*,*formal*] *ming*$^{21}_{[YUE]}$ / mei$_{[JP]}$	motorcycle [*vehicle*] 架 *gaa*$^{33}_{[YUE]}$ / 台 dai$_{[JP]}$	cassette 餅 *beng*$^{35}_{[flat,thick]}$ [*YUE*] / 本 hon$_{[prolonged]}$ [*JP*]
boat 艘 [*smallboat*] *sau*$^{35}_{[YUE]}$ / soo$_{[JP]}$	daisy [*flower*] 朵 *doe*$^{35}_{[YUE]}$ / 輪 rin$_{[JP]}$	aeroplane 部 *bou*$^{35}_{[machine]}$ [*YUE*] / 機 ki$_{[airvehicle]}$ [*JP*]
toilet paper 卷 [*roll*] *gyun*$^{35}_{[YUE]}$ / maki$_{[JP]}$	paper [*flat*,*thin*] 張 *zoeng*$^{55}_{[YUE]}$ / 枚 mai$_{[JP]}$	piegon 隻 *zek*$^{33}_{[animal]}$ [*YUE*] / 羽 wa$_{[bird]}$ [*JP*]
tree branch 枝 [*prolonged*,*fixed*] *zi*$^{55}_{[YUE]}$ / eda$_{[JP]}$	magazine [*book*] 本 *bun*$^{35}_{[YUE]}$ / 冊 satsu$_{[JP]}$	umbrellas 把 *baa*$^{35}_{[handle]}$ [*YUE*] / 本 hon$_{[prolonged]}$ [*JP*]

Table: Comparison of Classifier System in Japanese and Cantonese

Design II

Entity	Cantonese	Japanese
cassette	餅 [*thick*,*flat*] *beng*^{35}	本 [*prolonged*] *hon*
coin	個 [*round*.*object*] *go*^{33}	枚 [*piece*] *mai*
t-shirt	件 [*clothing*] *gin*^{22}	枚 [*piece*] *mai*
aeroplane	部 [*machine*] *bou*^{22}	機 [*airplane*] *ki*
company	間 [*indoor*.*location*] *gaan*^{55}	社 [*company*] *sya*
pigeon	隻 [*animal*] *zek*^{33}	羽 [*bird*] *wa*
sword	把 [*handle*] *baa*^{35}	本 [*prolonged*] *hon*
snake	條 [*prolonged*] *tiu*^{21}	匹 [*animal*] *hiki*
desk	張 [*flat*] *zoeng*^{55}	台 [*furniture*] *dai*
umbrella	把 [*handle*] *baa*^{35}	本 [*prolonged*] *hon*
article	篇 [*writing*] *pin*^{55}	本 [*prolonged*] *hon*
hat	頂 [*headwear*] *deng*^{35}	つ [*default*] *tsu*

Table: Experimental Items

- 4 conditions: Cantonese/Japanese CL – Cantonese–Japanese noun ⇒ 48 items (divided into 4 groups)
- Acceptability Judgement Test: 1–7 Likert Scale

Design III

(6) a. カセット-テープ-が 4 本/ 餅 見つか-った
kasetto-teepu-ga yon ho/ *beng*^{35} mitsuka-tta
cassette-tape-NOM four CL.JP$_{[+prolonged]}$/ CL.YUE$_{[+thick+flat]}$ find-PST
'I found four cassettes.'

b. 我 搵 到 四 餅/ 本
ngo^{23} *wan*^{35} *dou*^{35} *sei*^{33} *beng*^{35}/ hon
1 find DOU four CL.YUE$_{[+thick+flat]}$/ CL.JP$_{[+prolonged]}$
錄音帶
luk^{22}*jam*^{55}*-daai*^{35}
audio-tape
'I found four cassettes.' [Cantonese–Japanese]

Design IV

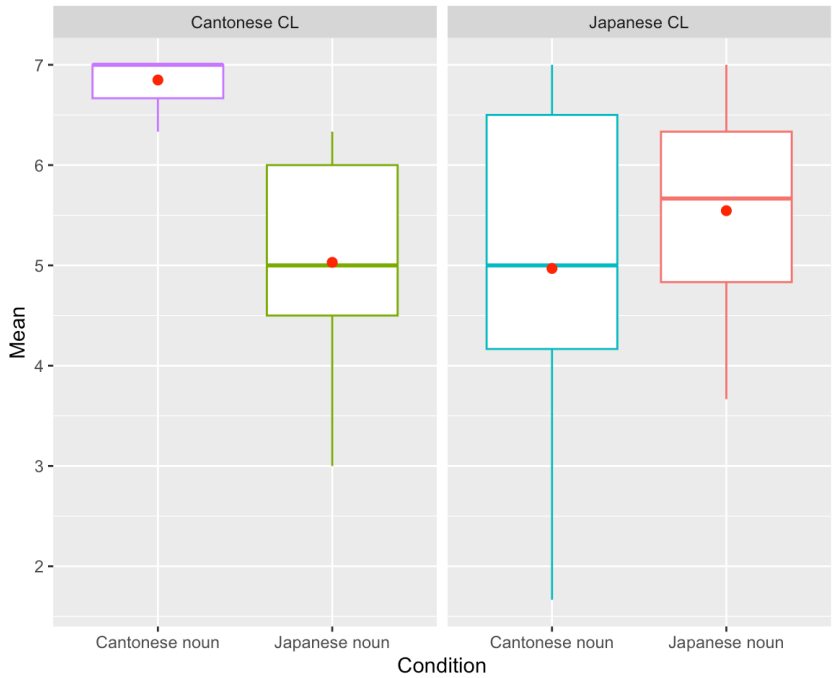
- Implemented on *PCIbex Farm*
- Before the trial: consent and instruction
- After the trial: anonymous demographic questionnaire

Participants

- A relatively small bilingual population, formed by the immigration wave in recent years
- 11 bilingual speakers (11 female; aged from 29;9 to 40;1) from Hong Kong
- Currently live or have been living in Japan for a period of time (around 5-8 years), work or daily language is Japanese
- recruited through personal networks and outreach

Result I

	YUE noun	JP noun
YUE CL	6.848	**5.030**
JP CL	**4.970**	5.545



Major takeaways from the result:

- The mixing variations in the classifier are prone to being acceptable
- As expected, the monolingual classifier conditions are still rated higher
- The relatively lower rating in the Japanese monolingual classifier condition may be due to unfamiliarity with the classifier in Japanese, which is also reflected in the mixing condition between a Japanese classifier and a Cantonese noun

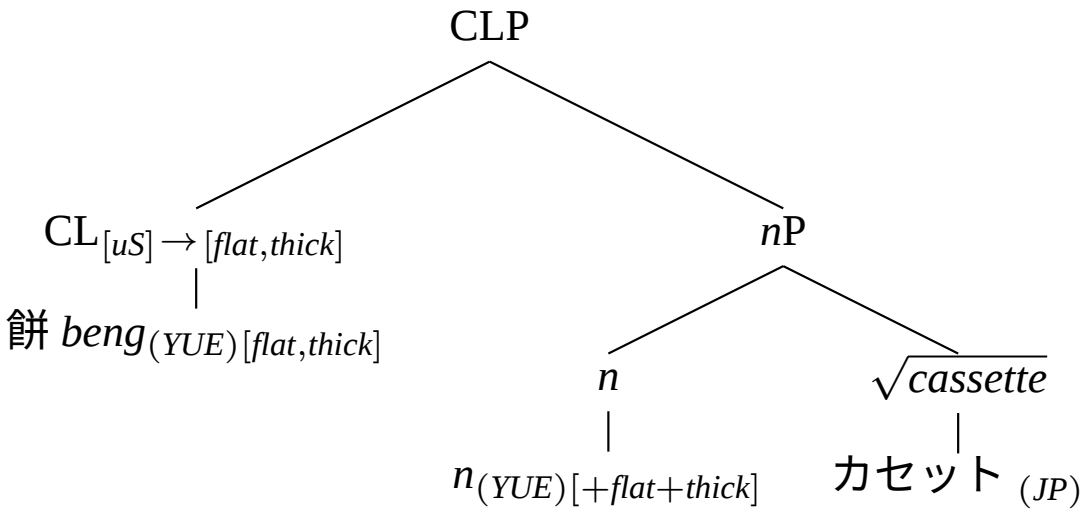
Discussion

- The *n*-approach captures the mixing variations in classifier use in the Cantonese–Japanese bilingual variety
  - Classifier Selectional Information: at *n* as a [feature]
  - *n* and the classifier at higher functional layer: Agree
  - Subset Principle for phonological realisation

Morphosyntactic Analysis I

(7) a. $n_{[+flat+thick]}$ b. $n_{[+prolonged]}$

(8)



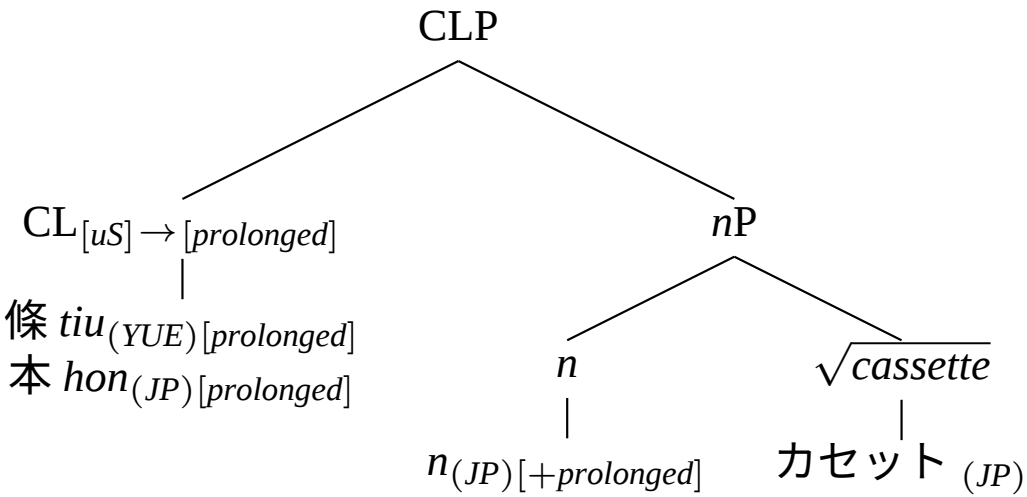
(9) [C:[uS:_][*n*[iS:+flat+thick] $\sqrt{cassette}$]] → [CL[flat,thick][*n*[iS:+flat+thick] $\sqrt{cassette}$]]

(10) a. $\sqrt{Root}^{126}$ ↔ 錄音帶 [YUE], カセット [JP] 'cassette'

b. $CL_{[flat,thick]}$ ↔ 餅 $beng_{[YUE]}$
$CL_{[flat]}$ ↔ 張 $zoeng_{[YUE]}$

Morphosyntactic Analysis II

(11)



(12) [C:[uS:_][*n*[iS:+prolonged] $\sqrt{cassette}$]] → [CL[prolonged][*n*[iS:prolonged] $\sqrt{cassette}$]]

(13) $CL_{[prolonged]} \leftrightarrow$ 條 $tiu_{[YUE]}$
$CL_{[prolonged]} \leftrightarrow$ 本 $hon_{[JP]}$

Licensing condition I

1. Semantic licensing conditions "are encoded in the Encyclopedia as conditions on the semantic interpretation of a $\sqrt{Root}$ in a context" (Kramer, 2015, 51)
   - The $\sqrt{Root}$ and the *n* have to match the encoded condition in the Encyclopedia, otherwise, the derivation will crash
   - Note that: "**crash**" in the Encyclopedia does not mean completely uninterpretable (other linguistic cues can help "repair" the interpretation)
2. Arbitrary licensing condition are "listed whether a certain *n* is included in the exponence of the $\sqrt{Root}$" (Kramer, 2015, 54) similar to a spell-out rule

Licensing condition II

Type	**Licensing condition**
Rigid choice interpretable e.g. functional	Semantic licensing condition:(Encyclopedia) only one *n* and with a specific feature/ a feature bundle combines with a $\sqrt{Root}$ to get interpretation e.g. [*n*[+book][$\sqrt{book}$]] = 'a book' → otherwise, it will crash in the Encyclopedia → get odd/infelicitous reading (or worst: uninterpretable)
Flexible choice interpretable e.g. shape	Semantic licensing condition:(Encyclopedia) multiple *n*s with a feature/ a feature bundle can match with a $\sqrt{Root}$ to get interpretation e.g. [*n*[+flat][$\sqrt{paper}$]] = 'a flat paper' e.g. [*n*[+roll][$\sqrt{paper}$]] = 'a roll of paper' → get different interpretations as long as the licensing condition fits with the entry in the Encyclopedia
Arbitrary choice uninterpretable	Arbitrary licensing condition: limited roots list their licensing conditions in the phonological term e.g. $\sqrt{mouth} \leftrightarrow$ [□] / *n*[+handle]

One lexical entry for each entity

Lexicalist approaches:

- assuming the information is stored at the lexical item
- involving duplications when multiple languages are acquired
- e.g. Japanese カセット 'cassette' needs to be copied to Cantonese, and add the new classifier information 餅 $beng_{(YUE)[flat,thick]}$ and 條 $tiu_{(YUE)[prolonged]}$
- The more entities or the more languages one speaks, the effect is multiplied

n approach:

- assuming the $\sqrt{Root}$ is shared in the same computational system
- any variations come from functional layers
- Capturing any language internal, language-mixing variations: the $\sqrt{Root}$ remains the same, only the *n* is different
- e.g. Japanese カセット 'cassette' is just a realisation of a $\sqrt{Root}^{126}$, the variations are from $n_{[+flat+thick]}$ $n_{[+prolonged]}$
- No reduplication needed: one $\sqrt{Root}$ for each entity, and functional layers remains the same

Conclusion

- The *n*-approach assumes a $\sqrt{Root}$ is syntactically and semantically void
- Semantic selectional information associated with classifier use is not at $\sqrt{Root}$ but *n*
- Contrast to the lexicalist approaches, redundant reduplication is avoided

Thank you!
多謝！
ありがとうございます！

References I

Acquaviva, P. (2009). Roots and lexicality in distributed morphology. In A. Galani, D. Redinger, & N. Yeo (Eds.), *YPL2-Issue 10 (May 2009) Special Issue-York-Essex Morphology Meeting (YEMM)*.

Aikhenvald, A. Y. (2025). *A guide to gender and classifiers*. Oxford University Press.

Aikhenvald, A. Y., & Mihas, E. I. (2019). *Genders and classifiers: A cross-linguistic typology*. Oxford University Press.

Alexiadou, A., & Lohndal, T. (2018). Units of language mixing: A cross-linguistic perspective. *Frontiers in psychology*, *9*, 1719.

Arad, M. (2003). Locality constraints on the interpretation of roots: The case of hebrew denominal verbs. *Natural language & linguistic theory*, *21*(4), 737–778.

Embick, D. (2015). *The morpheme: A theoretical introduction* (Vol. 31). Berlin, Germany: Walter de Gruyter GmbH & Co KG.

Grimstad, M. B., Lohndal, T., & Åfarli, T. A. (2015). Language mixing and exoskeletal theory: A case study of word-internal mixing in american norwegian. *Nordlyd*, *41*(2), 213.

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building* (Vol. 20). MIT Press.

References II

Harley, H. (2014). On the identity of roots. *Theoretical linguistics*, *40*(3-4), 225–276.

Kramer, R. (2015). *The morphosyntax of gender* (No. 58). Oxford University Press.

López, L. (2020). *Bilingual grammar: Toward an integrated model*. Cambridge, UK: Cambridge University Press.

Marantz, A. (2001). Words and things. *handout, MIT*.

Riksem, B. R. (2018). *Language mixing in american norwegian noun phrases. an exoskeletal analysis of synchronic and diachronic patterns.* (Unpublished doctoral dissertation). NTNU.