A construction-based approach to Cantonese classifiers

Francis <u>Bond</u> (凡土・フランシス) and Joanna <u>Sio</u> (蕭月嫦)

Computational Linguistics Lab (凡土研)
Department of Asian Studies
Palacký University, Olomouc

⟨bond@ieee.org, neosome@gmail.com⟩

HPSG 2024

Faculty of Arts

Overview

- A construction-based approach for NPs with sortal classifiers
- Uses only one specifier
- Accounts for information structure
- Accounts for 'missing' constituents
- Available at ⟨https://github.com/neosome/yue⟩.

Cantonese

- A variety of Yue, belongs to the Sinitic branch of the Sino-Tibetan language family.
- Cantonese is spoken in Guangdong China, and the two Special Administrative Regions, Hong Kong and Macao, as well as in diaspora communities (e.g., Singapore, Malaysia, Australia, the United Kingdom and North America).
- Over 82 million Cantonese native language speakers.
- Not widely written (instead Mandarin Chinese is used) or taught in schools
- Used widely in Opera, Films, Music (CantoPop)

All examples are in Cantonese unless otherwise specified.

An NP with classifier (DXCN)

(1)　明恩　　　食咗　　呢　一　個　蘋果。
　　　Ming4jan1 sik6-zo2 nei1 jat1 go3 ping4gwo2
　　　Ming-Jan　　eat-PERF　this　one　CL　apple

　　'Ming-Jan ate this apple.'
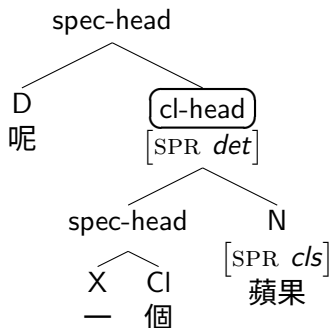
Possible Structures I

(2) Cantonese (yue)

  a. D-(X)-C-N
    明恩　　　食咗　　呢　(一) 個　蘋果。
    Ming4jan1 sik6-zo2 nei1 jat1 go3 ping4gwo2
    Ming-Jan　eat-PERF　this　one　CL　apple

    'Ming-Jan ate this apple.'

  b. X-C-N
    明恩　　　食咗　　一　個　蘋果。
    Ming4jan1 sik6-zo2 jat1 go3 ping4gwo2
    Ming-Jan　eat-PERF　one　CL　apple

    'Ming-Jan ate one apple.'

Possible Structures II

c. C-N
明恩　　　食咗　　個　蘋果。
Ming4jan1 sik6-zo2 go3 ping4gwo2
Ming-Jan　eat-PERF　CL　apple

'Ming-Jan ate an/the apple.'

d. N
明恩　　　食咗　　蘋果。
Ming4jan1 sik6-zo2 ping4gwo2
Ming-Jan　ate-PERF　apple

'Ming-Jan ate an apple/apples.'

Definiteness

In Chinese, in general, only definite NPs can appear in the subject or topic position in a sentence (Li and Thompson, 1989). We model this as follows:

Type	Example	cog-st	Definiteness
d-(x)-c-n	呢 (一) 個蘋果	fam-or-more	Definite
x-c-n	一個蘋果	type-id	Indefinite
c-n	個蘋果	fam-or-less	In/Definite
n	蘋果	type-id	Indefinite (or Generic)

Cognitive Status

Definiteness is the grammatical encoding of the concept of identifiability (Chen, 2004).The choice of referring expressions is determined by accessibility (understood as a property of representation of memory). We use the model of Gundel et al. (1993) expressed as types by Borthen and Haugereid (2005) and refined by Bender and Goss-Grubbs (2008):
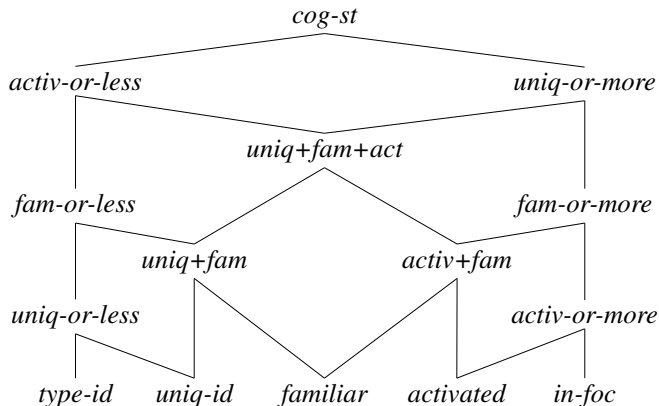
How many specifiers?

- Ng (1997) created an analysis of classifiers using two specifiers
  - nouns have $\left[\text{SPEC} \ \langle \textit{cls, det} \rangle\right]$
  - head-specifier discharges the first and links a reduced list
- instead of modifying head-specifier (head-spec) we create a head-classifier construction (head-cl) which requires another specifier after consuming the classifier
- the classifier-construction is the locus of the unusual syntax
- empirical data from a wide range of languages does not require two specifiers for an adequate description of noun phrases
- but we have not (yet) found compelling evidence from Cantonese to prefer one analysis over the other

Classifier Head Rule (*cl-head*: ver 1)

$$
\begin{bmatrix}
\textit{cl-head-phrase} \\[2pt]
\text{SYN} \quad
\begin{bmatrix}
\text{VAL.SPR} \quad \left\langle \begin{bmatrix} \text{HEAD } \textit{det} \end{bmatrix} \right\rangle \\
\text{SEM.INDEX} \quad \boxed{1}
\end{bmatrix} \\[14pt]
\text{HD} \quad
\begin{bmatrix}
\text{VAL.SPR} \quad \left\langle \boxed{2} \begin{bmatrix} \text{HEAD } \textit{cls} \end{bmatrix} \right\rangle \\
\text{SEM.INDEX} \quad \boxed{1}
\end{bmatrix} \\[14pt]
\text{NHD} \; \boxed{2}
\begin{bmatrix}
\text{VAL.SPR} \quad \left\langle \, \right\rangle \\
\text{SEM.XARG} \quad \boxed{1}
\end{bmatrix}
\end{bmatrix}
$$

Very similar to head-specifier, but requires a new specifier!

Noun

$$
\begin{bmatrix}
\textit{noun-lex} \\[2pt]
\text{SYN} \begin{bmatrix} \text{HEAD} & \textit{noun} \\[4pt] \text{VAL.SPR} & \left\langle \begin{bmatrix} \text{SYN.HEAD} & \textit{cls,} \\ \text{OPT} & + \end{bmatrix} \right\rangle \end{bmatrix} \\[16pt]
\text{SEM} \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{INDEX} & \boxed{1} \end{bmatrix} \\[4pt] \text{RELS} & \left\langle \begin{bmatrix} \text{ARG0} & \boxed{1} \ \textit{ref-ind} \end{bmatrix} \right\rangle \end{bmatrix}
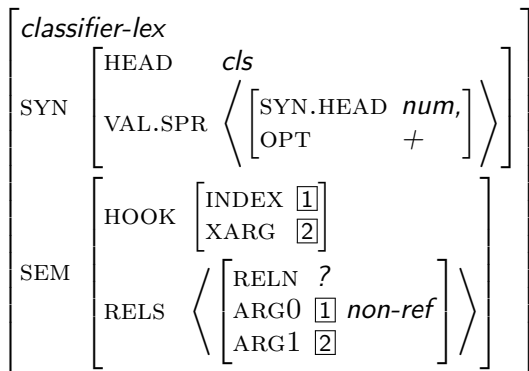\end{bmatrix}
$$
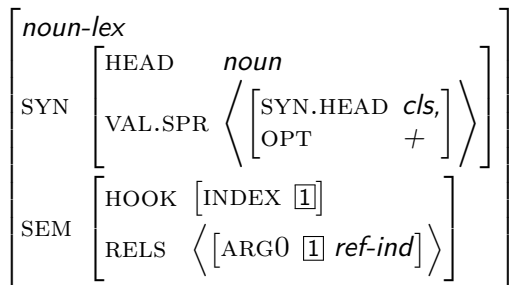
The Cantonese *noun-lex* sets its specifier to be a classifier, not a determiner. *[D-N] is ungrammatical in Cantonese.

Classifier

$$
\begin{bmatrix}
\textit{classifier-lex} \\
\text{SYN} \begin{bmatrix} \text{HEAD} & \textit{cls} \\ \text{VAL.SPR} & \left\langle \begin{bmatrix} \text{SYN.HEAD} & \textit{num}, \\ \text{OPT} & + \end{bmatrix} \right\rangle \end{bmatrix} \\
\text{SEM} \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{INDEX} & \boxed{1} \\ \text{XARG} & \boxed{2} \end{bmatrix} \\ \text{RELS} & \left\langle \begin{bmatrix} \text{RELN} & \textit{?} \\ \text{ARG0} & \boxed{1}\ \textit{non-ref} \\ \text{ARG1} & \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

Introduces a non-referential predicate (to model the semantic contribution).

Other approaches

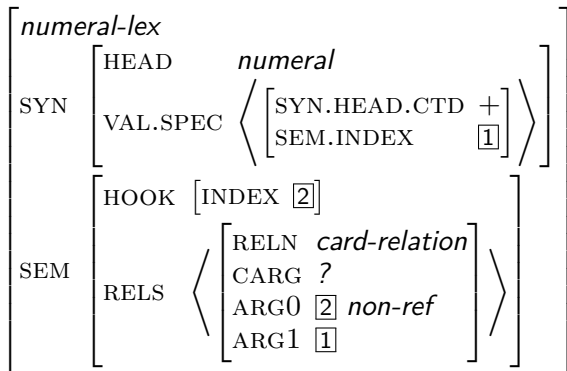- Co-index the classifier and noun (Bender and Siegel, 2004)
- Add the classifer as a field in a generalized measurement relationship, with the dimension set by the classifier and the cardinality by the number (郡司, 2005)

$$\begin{bmatrix} \text{RELN} & measure \\ \text{ARG1} & x_1 \\ \text{NUM} & 1 \\ \text{DIMENSION} & 個\_x \end{bmatrix}$$

- For now we will keep them separate, and as we analyze more structures we hope to find evidence one way or the other.

Numeral

$$
\begin{bmatrix}
\textit{numeral-lex} \\[4pt]
\text{SYN} \begin{bmatrix} \text{HEAD} & \textit{numeral} \\[4pt] \text{VAL.SPEC} & \left\langle \begin{bmatrix} \text{SYN.HEAD.CTD} & + \\ \text{SEM.INDEX} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \\[20pt]
\text{SEM} \begin{bmatrix} \text{HOOK} & \begin{bmatrix} \text{INDEX} & \boxed{2} \end{bmatrix} \\[4pt] \text{RELS} & \left\langle \begin{bmatrix} \text{RELN} & \textit{card-relation} \\ \text{CARG} & \textit{?} \\ \text{ARG0} & \boxed{2}\ \textit{non-ref} \\ \text{ARG1} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

Introduces a non-referential predicate (to model the cardinality).
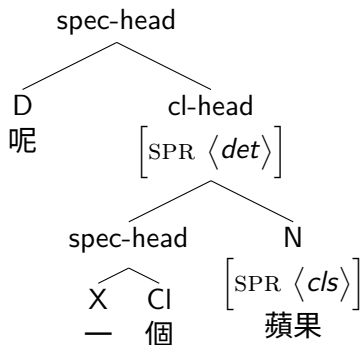
Demonstrative

$$
\begin{bmatrix}
\textit{det-lex} \\
\text{SYN} \begin{bmatrix}
\text{HEAD} & \textit{det} \\
\text{VAL.SPEC} & \left\langle \left[\text{LOCAL.SEM.HOOK.INDEX } \boxed{1} \right] \right\rangle
\end{bmatrix} \\
\text{SEM} \begin{bmatrix}
\text{HOOK} & \left[\text{INDEX } \boxed{1} \left[\text{COG-ST } \textit{fam-or-more} \right] \right] \\
\text{RELS} & \left\langle \left[\text{ARG0 } \boxed{1} \textit{ref-ind} \right] \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

A demonstrative constrains the index of the noun it specifies to be
fam-or-more.

Put it together (DXCN)

(3) 明恩食咗呢一個蘋果。

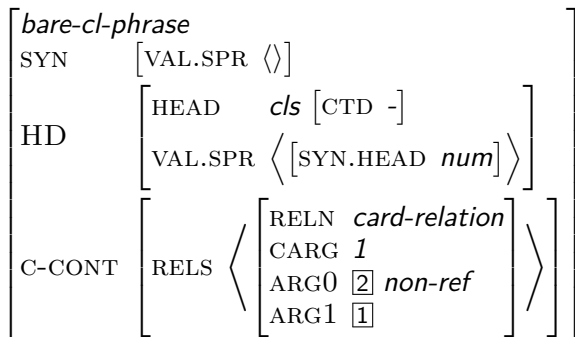呢 $\_q(x_1, h_2, h_3)$; card(e_4, x_1 '1'), 個 $\_x(e_5, x_1)$; 蘋果 $\_n(x_1)$

x_1 *fam-or-more*
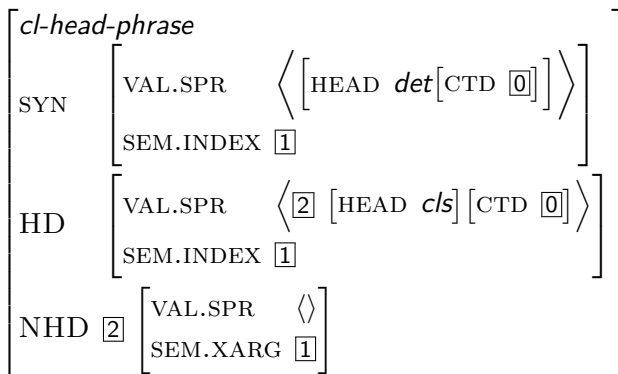
Bare classifier rule

This non-branching rule takes a classifier, and creates a classifier phrase.

$$
\begin{bmatrix}
\textit{bare-cl-phrase} \\
\text{SYN} \quad \begin{bmatrix} \text{VAL.SPR } \langle\rangle \end{bmatrix} \\
\text{HD} \quad \begin{bmatrix} \text{HEAD} \quad \textit{cls} \begin{bmatrix} \text{CTD -} \end{bmatrix} \\ \text{VAL.SPR } \left\langle \begin{bmatrix} \text{SYN.HEAD } \textit{num} \end{bmatrix} \right\rangle \end{bmatrix} \\
\text{C-CONT} \quad \begin{bmatrix} \text{RELS} \left\langle \begin{bmatrix} \text{RELN } \textit{card-relation} \\ \text{CARG } \textit{1} \\ \text{ARG0 } \boxed{2} \textit{ non-ref} \\ \text{ARG1 } \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

If there is no number, the interpretation is that there is only one referent. The feature CTD marks whether something has been explicitly counted or not. C-N cannot be the answer to the question 'how many?'.
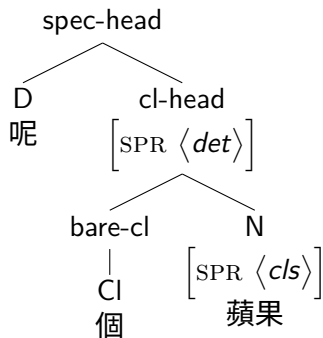
Classifier Head Rule (*cl-head*: ver 2)

$$
\begin{bmatrix}
\textit{cl-head-phrase} \\
\text{SYN} \quad \begin{bmatrix} \text{VAL.SPR} & \left\langle \begin{bmatrix} \text{HEAD} \ \textit{det}\begin{bmatrix}\text{CTD} \ \boxed{0}\end{bmatrix} \end{bmatrix} \right\rangle \\ \text{SEM.INDEX} \ \boxed{1} \end{bmatrix} \\
\text{HD} \quad \begin{bmatrix} \text{VAL.SPR} & \left\langle \boxed{2} \ \begin{bmatrix}\text{HEAD} \ \textit{cls}\end{bmatrix}\begin{bmatrix}\text{CTD} \ \boxed{0}\end{bmatrix} \right\rangle \\ \text{SEM.INDEX} \ \boxed{1} \end{bmatrix} \\
\text{NHD} \ \boxed{2} \ \begin{bmatrix} \text{VAL.SPR} & \langle \rangle \\ \text{SEM.XARG} & \boxed{1} \end{bmatrix}
\end{bmatrix}
$$

The classifier head rule links the value of CTD.
Determiners ignore it, bare NP rules look at it.
(We expect modifiers consider CTD too)

Put it together (DCN)

(4) 明恩食咗呢個蘋果。

```
                    spec-head
                   /         \
                  D          cl-head
                  呢         [SPR ⟨det⟩]
                            /          \
                      bare-cl           N
                         |          [SPR ⟨cls⟩]
                        Cl              蘋果
                        個
```

呢 $\_q(x_1, h_2, h_3)$; card(e_4, x_1 '1'), 個 $\_x(e_5, x_1)$; 蘋果 $\_n(x_1)$

x_1 *fam-or-more*

Bare noun phrase rule (N, XCN)

$$
\begin{bmatrix}
\textit{bare-np-phrase} \\
\text{SYN} \quad
\begin{bmatrix}
\text{VAL.SPR} & \langle\rangle \\
\text{SEM.INDEX} & \boxed{1}\ \begin{bmatrix}\text{COG-ST} & \textit{type-id}\end{bmatrix}
\end{bmatrix} \\
\text{HD} \quad
\begin{bmatrix}
\text{HEAD} & \textit{noun} \\
\text{VAL.SPR} & \left\langle \begin{bmatrix}\text{HEAD} & \textit{cls-or-det}\ \begin{bmatrix}\text{CTD} & +\end{bmatrix}\end{bmatrix} \right\rangle \\
\text{SEM.INDEX} & \boxed{1}
\end{bmatrix} \\
\text{C-CONT} \quad
\begin{bmatrix}
\text{RELS} & \left\langle \begin{bmatrix}\text{RELN} & \textit{exist\_q} \\ \text{ARG0} & \boxed{1}\end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

Only use if it has no classifier $\begin{bmatrix}\text{SPR} & \textit{cls}\end{bmatrix}$ or the classifier is explicitly counted $\begin{bmatrix}\text{SPR} & \textit{det}\ \begin{bmatrix}\text{CTD} & +\end{bmatrix}\end{bmatrix}$. Make it indefinite.

Put it together (XCN)

(5) 明恩食咗一個蘋果。

bare-NP
|
cl-head
$$\left[\text{SPR } \left\langle \textit{det } \begin{bmatrix} \text{CTD } + \end{bmatrix} \right\rangle \right]$$

spec-head
$$\left[\text{SPR } \left\langle \textit{cls } \begin{bmatrix} \text{CTD } + \end{bmatrix} \right\rangle \right]$$
N
蘋果

X   Cl
一   個
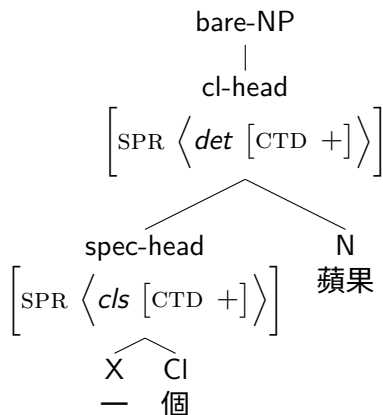
exist_q(x_1, h_2, h_3); card(e_4, x_1 '1'), 個 _x(e_5, x_1); 蘋果 _n(x_1)

x_1 *type-id*

Put it together (N)

(6) 明恩食咗蘋果。

$$\text{bare-NP}$$
$$|$$
$$\text{N}$$
$$\left[\text{SPR } \textit{cls} \langle \text{CTD } \pm \rangle \right]$$
$$\text{蘋果}$$
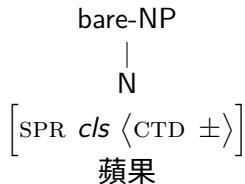
exist_q(x_1, h_2, h_3); 蘋果 _n(x_1)
x_1 *type-id*
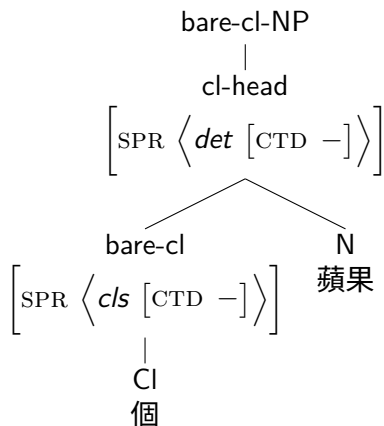
Bare noun phrase rule with bare classifier phrase (CN)

$$
\begin{bmatrix}
\textit{bare-cl-np-phrase} \\
\text{SYN} \quad
\begin{bmatrix}
\text{VAL.SPR} & \langle \rangle \\
\text{SEM.INDEX} & \boxed{1} \; [\text{COG-ST} \; \textit{fam-or-less}]
\end{bmatrix} \\
\text{HD} \quad
\begin{bmatrix}
\text{HEAD} & \textit{noun} \\
\text{VAL.SPR} & \left\langle \begin{bmatrix} \text{HEAD} \; \textit{det} \; [\text{CTD} \; \text{-}] \end{bmatrix} \right\rangle \\
\text{SEM.INDEX} & \boxed{1}
\end{bmatrix} \\
\text{C-CONT} \quad
\begin{bmatrix}
\text{RELS} & \left\langle \begin{bmatrix} \text{RELN} & \textit{exist\_q} \\ \text{ARG0} & \boxed{1} \end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

Only use if the classifier is not explicitly counted $\begin{bmatrix} \text{SPR} \; \textit{det} \; [\text{CTD} \; -] \end{bmatrix}$.

Make it in/definite.

Put it together (CN)

(7) 明恩食咗個蘋果。

bare-cl-NP
|
cl-head
$$\left[\text{SPR} \ \left\langle \textit{det} \ \left[\text{CTD} \ - \right] \right\rangle \right]$$

bare-cl
$$\left[\text{SPR} \ \left\langle \textit{cls} \ \left[\text{CTD} \ - \right] \right\rangle \right]$$
|
Cl
個

N
蘋果

exist_q(x_1, h_2, h_3); card(e_4, x_1 '1'), 個 _x(e_5, x_1); 蘋果 _n(x_1)

x_1 *fam-or-less*

Faculty
of Arts

Summary

- A construction-based approach for NPs with sortal classifiers that uses only one specifier
  cl-head construction
- Accounts for 'missing' constituents, using
  bare-cl, bare-np, bare-cl-np
- Accounts for information structure
  uses the head feature CTD to share information
- Available at ⟨https://github.com/neosome/yue⟩.

ToDo

We would like to look at:

- Other classifier types
  measure, kind, portion, container
- Other things in the numeral position
  quantifiers, interrogatives
- Other things nearby (half, a bit more, …)
- Reduplicated classifiers
- Modification within and of the NP
- Model the definiteness restriction on subjects
- Other Chineses (Mandarin, Min, Hakka …)
- Other classifier languages (Japanese, Indonesian, Korean, …)

And finally analyze and treebank a corpus

Simple Modification is ok

(8) 一　　隻　　　大　　狗
    jat1 zek3　daai6　gau2
    one　CL.unit　big　　dog

    'One big dog' [yue]

(9) 一　　大　　　隻　　　狗
    jat1 daai6 zek3　gau2
    one　big　　CL.unit　dog

    'One big dog' [yue]

The latter is marked, it means something like one big salient dog

But not with intensificaiton

(10) * 一　　好　　大　　隻　　狗
　　　jat1 hou2 daai6 zek3 gau2
　　　one　very　big　CL.unit dog

　　'One very big dog' [yue]

(11) * 一　　大　　大　　隻　　狗
　　　jat1 daai6 daai6 zek3 gau2
　　　one　big　big　CL.unit dog

　　'One big big dog' [yue]

The meaning changes for container/measure CL

(12) 一　大　　碗　　蘋果
　　　jat1 daai6 wun2　ping4 gwo2
　　　one　big　CL.bowl apple

　　　'A big bowl of apples.'

(13) 一　碗　　大　　蘋果
　　　jat1 wun2　daai6 ping4 gwo2
　　　one　CL.bowl big　　apple

　　　'A bowl of big apples.'

Kind classifiers cannot be modified

(14) 一　種　　大　　動物
　　　jat1 zung2 daai6 dung6
　　　one CL.kind large animal

'A kind of large animal.'

(15) *一　大　　種　　動物
　　　jat1 daai6 zung2 dung6
　　　one large CL.kind animal

'A kind of large animal.'/'A large kind of animal.'

Acknowledgments

- Like to thank the reviewers
- Like to thank Dan Flickinger, Emily Bender and Luis Morgado da Costa for discussion

References I

Emily M. Bender and David Goss-Grubbs. 2008. Semantic Representations of Syntactically Marked Discourse Status in Crosslinguistic Perspective. In *Proceedings of the 2008 Conference on Semantics in Text Processing*, pages 17–29. Association for Computational Linguistics.

Emily M. Bender and Melanie Siegel. 2004. Implementing the syntax of Japanese numeral classifiers. In *Proceedings of the IJC-NLP-2004*.

Kaja Borthen and Petter Haugereid. 2005. Representing Referential Properties of Nominals. *Research on Language and Computation*, 3(2-3):221–246.

Ping Chen. 2004. Identifiability and definiteness in Chinese. *Linguistics*, 42(6):1129–1184.

Jeanette K. Gundel, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, 69(2):274–307.

References II

Charles N Li and Sandra A Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. Univ of California Press.

Say Kiat Ng. 1997. A double specifier account of Chinese NPs using head-driven phrase structure grammar. Msc thesis, Department of Linguistics, University of Edinburgh.

隆男郡司. 2005. Measurement and quantification revisited. *Theoretical and applied linguistics at Kobe Shoin :* トークス, 8:21–36.