



UNIVERSITAS
INDONESIA

Veritas, Probitas, Iustitia

FAKULTAS
ILMU
PENGETAHUAN
BUDAYA

Part of Speech Tagging: Meninjau Ulang Kelas Kata dalam bahasa Indonesia

Totok Suhardijanto

Departemen Linguistik, FIB Universitas Indonesia

Workshop INACL, Binus University, 11-12 Juli 2017

Titik Mula

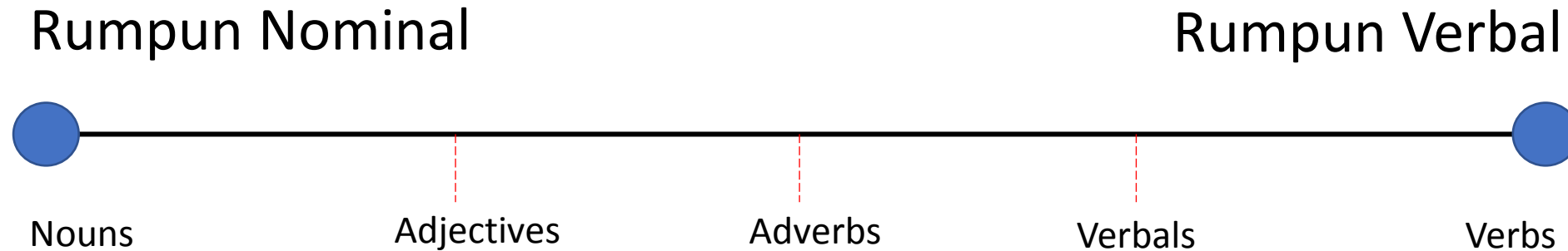
- Apa itu *part of speech* (POS)?
 - Lexically defined, natural classes of words that pattern together in important respects – phonologically, morphologically, syntactically
- Ahli bahasa (*linguist*) mendekati POS dengan pandangan:
 - 1) POS merupakan senarai kategori kata yang bersifat nyaris universal
 - 2) Apakah perbezaan kategorinya bersifat diskrit atau skalar? Apakah perbezaan nomina/verba itu merupakan kontinuum?
 - 3) Apakah ada perbezaan pada taraf akar (*root*), kata (*word*), dan kalimat (*syntactic unit*)?

POS bersifat nyaris universal

- Nomina, verba, dan adposisi ditemukan dalam hampir semua bahasa
- Namun, ada beberapa perbedaan detail dalam hal kelas kata pada banyak bahasa:
 - Bahasa Jepang punya adjektiva *i* and adjektiva *na*
 - Pada bahasa Swedia, untuk gender netral, tidak ada perbedaan di antara adverbial dan adjektiva:
 - a) *Försöket var **snabbt*** = 'Upaya itu sangat cepat';
 - b) *Försöket ändade **snabbt*** = 'Upaya itu berakhir cepat.'

Kategori bersifat diskrit atau skalar?

- Bahasa Inggris



POS bersifat tetap atau terikat konteks?

- bahasa Jawa

stem

layu

kata

mlayu

lumayu

playu

- bahasa Indonesia

tanpa objek

Dampaknya *sangat merugikan*.

dengan objek

Dampak korupsi *sangat merugikan* banyak orang.

Yāska, Ahli Tata Bahasa Sanskerta

- Dalam Nirukta, ditulis Abad 5 atau 6 SM,

नाम nāma – nomina (termasuk adjektiva)

आख्यात ākhyāta – verba

उपसर्ग upasarga – pra-verbal atau prefiks

निपात nipāta – partikel, kata tetap (mungkin preposisi)

Sībawayh, Ahli Tata Bahasa Arab

- Dalam *Alkitab*, ditulis Abad 8 M,

الأسم *ism* – nomina (di dalamnya termasuk adjektiva (*sifa*))

الفعل *fiil* – verba

الحرف *harf* – partikel

Dionysius Thrax (100 SM)

- 1) Nomina (*ónoma*): POS yang terinfleksi dalam kasus, menandai entitas konkret atau abstrak
- 2) Verba (*rhēma*): POS tanpa infleksi kasus, tetapi terinfleksi kala (*tense*), persona dan jumlah, menandai sebuah aktivitas atau proses yang dilakukan atau dialami
- 3) Partisiple (*metokhē*): POS yang berbagi fitur verba dan nomina
- 4) Article (*árthron*): POS yang dapat terdeklinasi, termasuk di dalamnya artikel definit, juga pronomina relatif

Dionysius Thrax (100 SM)

- 5) Pronomina (*antōnymía*): POS yang dapat menggantikan nomina dan menandai seseorang.
- 6) Preposisi (*próthesis*): POS yang ditempatkan sebelum kata lain dalam urutan dan dalam struktur kalimat
- 7) Adverb (*epírrhēma*): POS tanpa infleksi, sebagai pewatas (*modifier*) atau tambahan pada verba, adjektiva, klausa, kalimat, atau adverb lain.
- 8) konjungsi (*sýndesmos*): POS yang menyatukan wacana dan mengisi celah (*gap*) dalam interpretasinya.

Chomsky (1970)

- Dua fitur distingtif biner

POS	N	V
Nouns	+	-
Verbs	-	+
Adjectives	+	+
Adpositions	-	-

- Biasanya hanya digunakan untuk membedakan keempat kategori utama

Jackendoff (1977)

- Mirip dengan Chomsky tetapi berbeda fitur
 - +SUBJ, -OBJ = Noun
 - +SUBJ, +OBJ = Verb
 - SUBJ, -OBJ = Adjectives
 - SUBJ, +OBJ = Adposition
- Noun/Verb dan Adjective/Adposition adalah kelas natural bagi Jackendoff, berbeda dengan Chomsky

Baker (2003)

Noun adalah +N = 'mempunyai indeks referensial'

Verb adalah +V = 'mempunyai penciri'

Adjective adalah -N, -V

Adpositions bersifat fungsional (sistem yang berbeda)

Kelas Kata dalam bahasa Indonesia (1/4)

- Tata Bahasa Baku Bahasa Indonesia (1993)

- 1) verba
- 2) adjektiva
- 3) nomina
- 4) pronomina
- 5) numerelia
- 6) adverbialia
- 7) kata tugas

Kelas Kata dalam bahasa Indonesia (2/4)

- Harimurti Kridalaksana (1986)
 - 1) verba
 - 2) adjektiva
 - 3) nomina
 - 4) pronomina
 - 5) numerelia
 - 6) adverbialia
 - 7) interogativa
 - 8) demonstrativa
 - 9) artikula
 - 10) preposisi
 - 11) konjungsi
 - 12) kategori fatis
 - 13) interjeksi

Kelas Kata dalam bahasa Indonesia (3/4)

- Abdul Chaer (2006)

1) Kata benda

2) Kata ganti

3) Kata kerja

4) Kata sifat

5) Kata sapaan

6) Kata penunjuk

7) Kata bilangan

8) Kata penyangkal

9) Kata depan

10) Kata penghubung

11) Kata keterangan

12) Kata tanya

13) Kata seru

14) Kata sandang

15) Partikel penegas

Kelas Kata dalam bahasa Indonesia (4/4)

- Anton Moeliono (1967)
 - 1) rumpun nominal
 - 2) rumpun verbal
 - 3) rumpun partikel

Permasalahan: non-prototypicality

- *mengantar* dan *melamun* adalah verba prototipikal, *mengantuk* tidak

- | | | |
|--------------------------|-----|----------------------------|
| a. Husni sudah mengantuk | vs. | Husni *(sedang) mengantuk. |
| b. Rani sangat mengantuk | vs. | Rani sangat pintar. |
| c. mengantar-diantar | vs. | mengantuk-*dikantuk |
| d. melamun-melamunkan | vs. | mengantuk-?mengantukkan |

Permasalahan: kekurangan penciri gramatikal

- a. *Der* Seemann steht da vs. Si pelaut berdiri di sana.
- b. Ich sah *den* Seemann vs. Saya melihat si pelaut.
- c. 美しい景色 vs. pemandangan indah
- d. 彼の家を訪問しました。 vs. Kami mengunjungi rumahnya.

Tagset bahasa Indonesia

Adriani *et al* (2009)

- | | | |
|------------------------------------|----------------------------------|--------------------------------|
| 1) CC (coordinate conjunction) | 10) PR (common pronouns) | 19) , (comma) |
| 2) CD (cardinal numerals) | 11) PRP (personal pronouns) | 20) : (colon or ellipsis) |
| 3) FW (foreign words) | 12) RB (adverbs) | 21) ((opening parenthesis) |
| 4) IN (prepositions) | 13) SC (subordinate conjunction) | 22)) (closing parenthesis) |
| 5) JJ (adjectives) | 14) SYM (symbols) | 23) " (opening quotation mark) |
| 6) MD (modal or auxiliaries verbs) | 15) VB (verbs) | 24) " (closing quotation mark) |
| 7) NEG (negations) | 16) WDT (wh-determiners) | 25) . -- (dash) |
| 8) NN (common nouns) | 17) WH (WH) | |
| 9) NNP (proper nouns) | 18) . (sentence terminator) | |

Tagset bahasa Indonesia

Larasati *et al* (2011)

- | | | |
|---------------------------------|-----------------------------------|---------------------|
| 1) H (coordinating conjunction) | 8) G (negation) | 15) V (verb) |
| 2) C (numeral) | 9) N (noun) | 16) W (question) |
| 3) B (determiner) | 10) P (personal pronoun) | 17) O (copula) |
| 4) F (foreign word) | 11) D (adverb) | 18) X (unknown) |
| 5) R (preposition) | 12) T (particle) | 19) Z (punctuation) |
| 6) A (adjective) | 13) S (subordinating conjunction) | |
| 7) M (modal) | 14) I (interjection) | |

Tagset bahasa Indonesia

Dinakaramani *et al* (2012)

- | | | |
|--------------------------------|---|----------------------------------|
| 1) CC Coordinating conjunction | 9) NEG Negation | 17) SC Subordinating conjunction |
| 2) CD Cardinal number | 10) NN Noun | 18) SYM Symbol |
| 3) OD Ordinal number | 11) NNP Proper noun | 19) UH Interjection |
| 4) DT Determiner / article | 12) NND Classifier, partitive, and measurement noun | 20) VB Verb |
| 5) FW Foreign word | 13) PR Demonstrative pronoun | 21) WH Question |
| 6) IN Preposition | 14) PRP Personal pronoun | 22) X Unknown |
| 7) JJ Adjective | 15) RB Adverb | 23) Z Punctuation |
| 8) MD Modal and auxiliary verb | 16) RP Particle | |

Tagset bahasa Indonesia

INACL: Ayu Purwantiari, Gunarso, Totok Suhardijanto (2017)

Category	Subcategory	Tagset	Example
Noun	Common Noun	NNO	pulau
	Proper Noun	NNP	Sumatera
Pronoun	Personal Pronoun	PRN	dia
	Relative Pronoun	PRR	yang
	Interogative	PRI	siapa
	Clitized Pronoun	PRK	-nya
Adjective	Adjective	ADJ	santun

Tagset bahasa Indonesia

INACL: Ayu Purwantiari, Gunarso, Totok Suhardijanto (2017)

Category	Subcategory	Tagset	Example
Verb	Intransitive Verb	VBI	menangis
	Transitive Verb	VBT	menangisi
	Passive Verb	VBP	ditangisi
	Linking Verb	VBL	adalah
	Existensial Verb	VBE	ada
Adverb	Adverb	ADV	sangat
	Aspectual and Tense Adverb	ADK	sedang
	Negation	NEG	tidak

Tagset bahasa Indonesia

INACL: Ayu Purwantiari, Gunarso, Totok Suhardijanto (2017)

Category	Subcategory	Tagset	Example
Conjunction	Coordinating Conj	CCN	dan
	Subordinating Conj	CSN	karena
Preposition	Preposition	PRP	di
Interjection	Interjection	INT	wah
Determiner	Quantifier	QUA	ekor
	Numeral	NUM	satu, 13
	Article	ART	si
Particle	Particle	PAR	pun

Tagset bahasa Indonesia

INACL: Ayu Purwantiari, Gunarso, Totok Suhardijanto (2017)

Category	Subcategory	Tagset	Example
Unit Symbol	Unit Symbol	UNS	cm
Currency	Currency	\$\$\$	Rp
Other symbol	Symbol	SYM	;

Prinsip POS tagset

1) Signifikan secara linguistik:

- Penentuan kelas/subkelasnya berdasarkan kepentingannya dalam konstruksi yang lebih besar: frasa, kalimat, dst
- Apakah perbedaan verba pasif vs. verba aktif fungsional?
- Apakah perlu perbedaan nomina terbilang vs. nomina tak terbilang?

2) Sempel dan praktis secara teknis:

- Penentuan kelas/subkelasnya berdasarkan kemudahan dalam penyusunan sistem pelabelan otomatis
- Apakah perlu perbedaan nomina tunggal vs. nomina jamak?
- Apakah perlu dibedakan nomina umum vs. nomina nama diri (*proper noun*)?

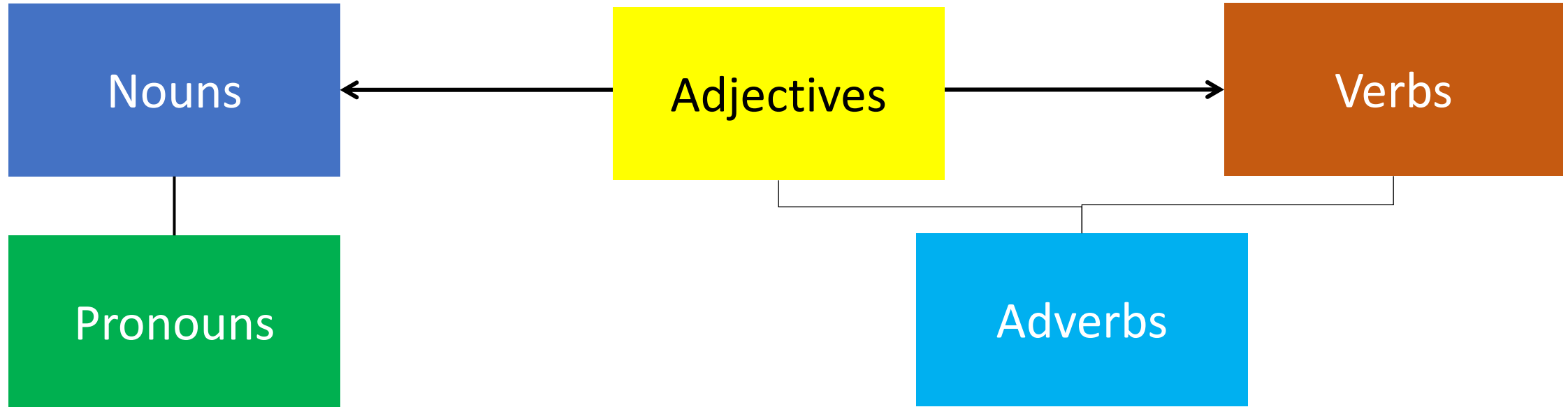
Konsep POS tagset

Kata Substantif

Kata Fungsi

Simbol/Lambang

Kata Substantif



Kata Fungsi & Simbol

Nouns

Verbs

Determiner

Preposition

Conjunction

Interjection

Particle

Unit Symbols

Currencies

Other Symbols

Contoh Pelabelan

Dengan pengobaran sentimen semacam ini, seolah bangsa Indonesia diingatkan

PPO NNO NNO PPO PRN SYM ADV NNO NNP VBT

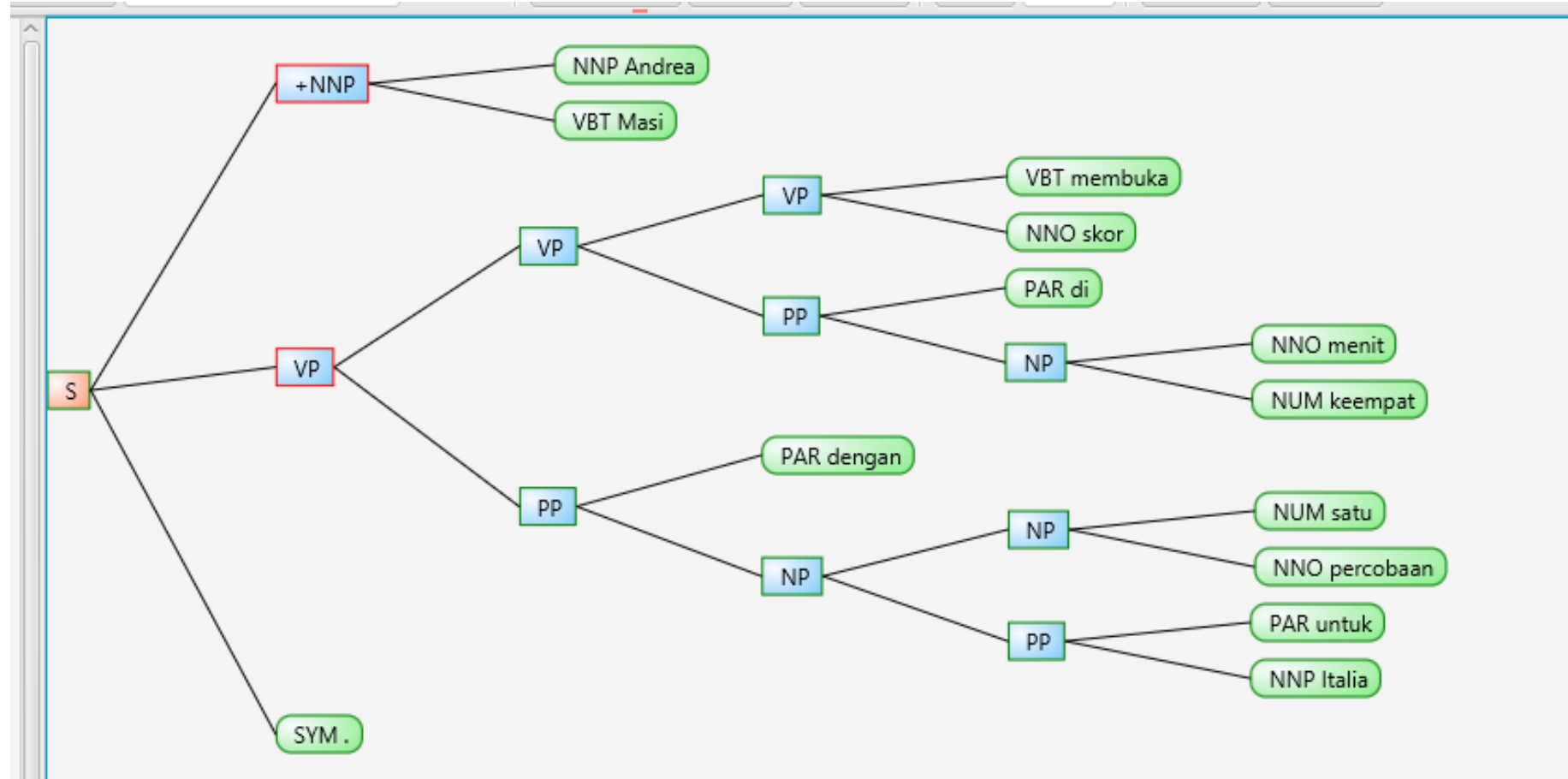
bahwa keturunan Cina bukanlah bagian dari mereka.

CSN NNO NNP NEG NNO PPO PRN SYM

>>>

Dengan PPO pengobaran NNO sentimen NNO semacam PPO ini PRN ,_SYM seolah ADV
bangsa NNO Indonesia NNP diingatkan VBT bahwa CSN keturunan NNO Cina NNP
bukanlah NEG bagian NNO dari PPO mereka PRN ,_SYM

Contoh Penerapan



Penutup

- Penambahan subkategori:
 - Subkategori *modal* pada *adverb*
 - Subkategori *punctuation* pada *other symbol*
- Ada perbedaan label pada level/taraf yang berbeda?
- Kelas kata apa saja yang signifikan dalam membedakan informasi sintaktis atau semantis?
- Dibutuhkan elaborasi baik dari segi teoretis dan praktis.