

Kategori, fungsi, dan peran sintaksis bahasa Indonesia dengan *PoS Tagging* berbasis *rule* dan *probability* (Categories, functions, and roles of Indonesian syntax using rule and probability based PoS Tagging)

Herpindo^{a1*}, Asri Wijayanti^{a2}, Irsyadi Shalima^{a3}, Ratih Ngestrini^{b4}^aFakultas Keguruan dan Ilmu Pendidikan, Universitas Tidar, Jl. Kapten Suparman No. 39, Potrobangsari, Kota Magelang Indonesia.^bDepartment of Computer Science, Utrecht University, Heidelberglaan 83584 CS Utrecht, The Netherlands.¹herpindo@untidar.ac.id; ²asriwijayanti@untidar.ac.id; ³irsyadi.shalima@untidar.ac.id; ⁴ratih.ngestrini@uu.nl

*Corresponding author: herpindo@untidar.ac.id

Sejarah Artikel

Diterima: 7 November 2021

Direvisi: 13 April 2022

Tersedia Daring: 29 April 2022

ABSTRAK

Penelitian ini dilatarbelakangi oleh kekurangan basis data gramatikal kategori khususnya pada verba yang belum lengkap yang menduduki fungsi predikat kopulatif, ekuatif, dan eksistensial dan belum adanya data *Pos Tagging* peran semantik agen dan pasien pada bahasa Indonesia. Diharapkan hasil penelitian ini dapat memprediksi kategori, fungsi, dan peran secara keseluruhan pada tiap kalimat. Penentuan kategori dilakukan dengan pembangunan model *probability-based PoS Tagging*, dan menghasilkan akurasi sebesar 92,53% dengan penambahan khusus tag verba (kopulatif, ekuatif, dan eksistensial) yang dapat menduduki posisi predikat dengan. Sementara itu, penentuan fungsi dan peran dilakukan dengan metode *rule-based*, yakni dengan pembuatan algoritma. Hasil akhir dari penelitian ini adalah kalimat baru dari *The corpus ind_mixed_2013 Leipzig Corpora* yang telah ditentukan kategori, fungsi, dan perannya untuk masing-masing kata. Pada penentuan peran agen, kedudukan agen pada kalimat yang diprediksi secara benar adalah sebanyak 82,58% dengan *row label* N dan PRON yang menduduki posisi agen bertipe nominative akusatif aktif, sedangkan pada peran pasien pada kalimat yang diprediksi secara benar adalah sebanyak 9,68% dengan *row label* N dan PRON dengan tipe pasif, ergative, dan anti-pasif. Temuan baru dalam penelitian ini adalah komponen verba (kopulatif, ekuatif dan eksistensial) yang dapat menduduki posisi predikat dalam secara fungsional dan potensi adanya tipologi ergative dengan diatesis ergative dan anti-pasif setelah diterapkan *rule* dan *probability based* dengan *Pos Tagging*.

Kata Kunci Fungsi, Kategori, PoS tagging, Probability, Rule

ABSTRACT

This research is motivated by the lack of a grammatical database of categories of incomplete verbs that occupy the functions of copulative, equative, existential predicates and the absence of *Pos Tagging* data on the semantic roles of agents and patients in Indonesian. It is hoped that the results of this study can predict the categories, functions, and roles in each sentence. The determination of the category is carried out by building a probability-based *PoS Tagging* model and produces an accuracy of 92.53% with the addition of special verb tags (copulative, equative, and existential) that can occupy the predicate position with. Meanwhile, functions and roles are determined using a rule-based method, namely by making algorithms. The final result of this research is a new sentence from *The corpus ind_mixed_2013 Leipzig Corpora*, which has determined the category, function, and role of each word. In determining the agent's role, the agent's position in the correctly predicted sentence is 82.58%, with the row label N and PRON occupying the position of the active accusative nominative agent. In comparison, the patient's role in the correctly predicted sentence is 9.68% with row label N and PRON with passive, ergative, and anti-passive types. The new finding in this study is that the verb components (copulative, equative, and existential) can occupy the predicate position in a functional way and the potential for an ergative typology with ergative and anti-passive diathesis after applying rules and probability-based with *Pos Tagging*.

Copyright©2022, Herpindo, Asri Wijayanti, Irsyadi S, Ratih Ngestrini
This is an open access article under the [CC-BY-3.0](https://creativecommons.org/licenses/by/3.0/) license

Keywords Function, Category, PoS tagging, Probability, Rule

How to Cite Herpindo, H., Wijayanti, A., Shalima, I., & Ngestrini, R. (2022). Kategori, fungsi, dan peran sintaksis bahasa Indonesia dengan *PoS Tagging* berbasis *rule* dan *probability*. *KEMBARA: Jurnal Keilmuan, Bahasa, Sastra dan Pengajarannya (e-Journal)*, 8(1), 51-65.
<https://doi.org/10.22219/kembara.v8i1.18602>

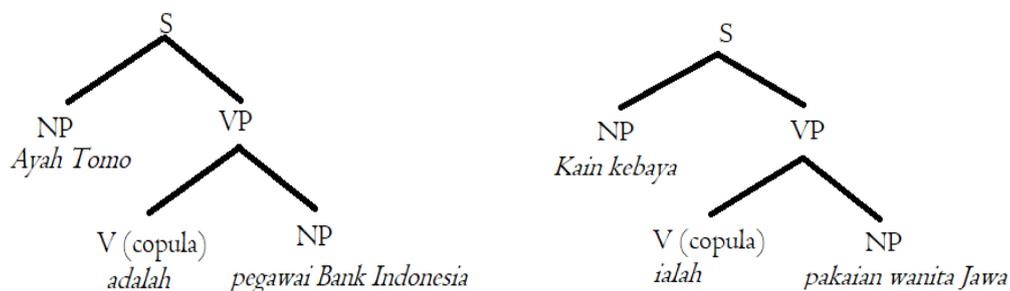


PENDAHULUAN

Kategori sintaksis yang memosisikan verba (V), subjek (S) sebagai *core*, peran agen (A) dan pasien (P) secara umum dalam sebuah klausa atau kalimat sangat sulit diidentifikasi, jika melihat fenomenanya dalam ragam bahasa Indonesia yang sangat kompleks. Hal ini didasarkan pada perbedaan sudut pandang linguistik dalam memetakan masalah disebabkan oleh data yang mereka peroleh. Verhaar (2006) membagi kategori menjadi nomina (N), numeralia (Num), Adjektiva (Adj), verba (V), konjungsi (Konj), keterangan (Ket), artikel (Art), kata seru dan preposisi (Prep); peran yang meliputi aktif, pasif, medial, benefaktif, objektif, agentif, instrumental, lokatif temporal, kausatif, kopulatif, eksistensial, progresif, posesif dll. Berbeda dengan Verhaar, Ramlan (1987) memberikan istilah pelengkap dalam fungsi dan tidak ada perbedaan dalam kategori hanya saja dalam tataran peran sintaksis ada beberapa tambahan seperti pemerolehan, hasil, mengalami, dikenal, terjumlah, peserta, keseringan, perbandingan, dan perkecualian. Kridalaksana (2008) juga memberikan tambahan pada tataran peran yang meliputi (arti) dari argument (peserta) struktur kalimat pada verba, sehingga peran atau fungsi semantis yang dimaksud adalah hubungan antar argumen atau predikator dalam proposisi.

Sebagai perbandingan, dalam kajian *Role Reference Grammar* (RRG) Rober (2001) menambahkan bentuk elemen predikat dan nonpredikat yang diduduki oleh frasa nomina yang berposisi sebagai argument predikat dan frasa adposisi sebagai nonargumen, sehingga dalam kalimat nonverbal nomina selanjutnya akan menjadi predikat. Tallerman (2015) memberikan tambahan *sub-classes* pada kategori verba atau predikat secara fungsional yang merupakan bagian dari hubungan antara kalimat. Teori *Lexical Functional Grammar* (LFG) yang dikembangkan pertama kali oleh Joan Bresnan dan Ronald Kaplan di tahun 1970-an dengan konsep *classical LFG* yang merupakan tata bahasa generative non-transformasional. Rentetan aturan dalam LFG memiliki informasi dalam struktur leksikal dan fungsional (Musgrave, 2013). Sejalan dengan hal tersebut Artawa (2020) memberikan penekanan terhadap leksikal dalam LFG dalam tataran kategorial kata. Berkaitan dengan LFG beberapa hal yang perlu dikemukakan adalah kategori leksikal dan kategori fungsional. Chomsky (dalam Falk, 2011) memberikan istilah *major lexical categories* terhadap N (Nomina), P (Preposisi), V (Verba) dan A (Adjektiva). Dalam LFG, fungsi lebih dideskripsikan dalam matematis yang berkorelasi dengan relasi grammatical (SUBJ, OBJ, OBL) yang menjadi fungsi gramatikal yang menekankan subjek (SUBJ) dan objek (OBJ) dalam hal ini Arka (2018) menyebutnya sebagai representasi sintaktik.

Melihat fenomena kategori (leksikal), fungsi (fungsional) dan peran (semantik) yang telah dikaji dalam kajian linguistik mikro, penelitian ini mencoba untuk merekonstruksi telaah struktural yang cenderung belum menyentuh ragam bahasa Indonesia yang jauh lebih kompleks. Hal ini bertujuan untuk menjawab beberapa hal terkait penentuan posisi leksikon yang dalam kasus bahasa Indonesia tidak selalu diduduki oleh verba secara konkret pada kategori sintaksis. Kasus seperti ini digambarkan oleh Sneddon et al (2010) sebagai kopula yang berkorespondensi dengan *linking verb* yang banyak muncul pada klausa nomina dalam bahasa Indonesia. Hal ini dapat dilihat pada diagram pohon Gambar I berikut.



Gambar I. Diagram Pohon Verba Kopulatif *adalah* dan *ialah* dalam Bahasa Indonesia

Walaupun dalam hal ini, Sneddon et al., (2010) berasumsi bahwa *adalah* dan *ialah* dalam bahasa Indonesia bukan merupakan verba, dalam ragam bahasa Indonesia *copula* ini sering muncul dan merupakan verba

nominalisasi (*nominalized verb*) dan secara fungsional merupakan argumen tak bersubjek. Verhaar (dalam [Ambiya, 2018](#)) memberikan gambaran mengenai kopula ini dengan menyebutnya sebagai predikat penyama, hal ini merupakan peran kopula yang berposisi sebagai predikat dengan distribusi pada Tabel 1.

Tabel 1
Distribusi Kopula Bahasa Indonesia

No	Relasi	Copula		Verba	
1	Klausa Kopula dengan Predikat FN	∅	kopula' (x, [pred'])	x = yang dilekati atribut, y = atributif	<i>Budi o guru</i>
		*adalah/*ialah/*merupakan	kopula' (x,y)	x, y = referen	<i>Budi adalah guru</i>
		*adalah/*ialah/*merupakan	kopula' (x,y)	x, y = referen	<i>Anak umar adalah mahasiswa Budi</i>
2	Klausa Kopula dengan Predikat FAdj	∅	kopula' (x, [pred'])	x= yang melekat pada atribut y=atribut	<i>Ani o cantik</i>
3	Klausa Kopulatif dengan Predikat FPrep	ada/berada	kopula' (x,y)	x, y = referen	<i>Budi (ada/berada) di rumah.</i>
		∅	kopula' (x, [pred'])	x= yang melekat pada atribut y=atribut	<i>Budi o di rumah.</i>

Berkaitan dengan fungsi dalam kalimat bahasa Indonesia, fungsi sintaksis akan diisi oleh berbagai kategori [Verhaar & Alip \(1996\)](#) dengan posisi S (subjek), P (predikat), Kom (komplemen), dan Ket (keterangan) yang akan diisi oleh kategori sintaksis, sehingga bentuk kata mempunyai fungsi sintaksis yang berbeda dalam kalimat. Fungsi memiliki kedudukan dalam struktur kalimat dengan unsur pengisi dalam kategori yang dapat dilihat pada Tabel 2.

Tabel 2
Relasi Fungsi dan Kategori dalam Bahasa Indonesia

No	Fungsi	Kategori
1	Subjek	nomina, frasa nomina, klausa nomina
2	Predikat	verba, ekuatif, dan ada (eksistensial)
3	Objek	nomina, frasa nomina, klausa nomina
4	Komplemen	(wajib hadir dan bukan nomina)
5	Adjungta	tidak wajib hadir (mana suka)

Fungsi adalah sekat-sekat kosong yang harus diisi oleh kategori kelas kata (*form filler*) kemudian pengisi kedua adalah peran semantis (*role filler*). Secara fungsional, terdapat tiga macam klausa yang predikatnya diduduki oleh verba, nomina, adjektiva dan preposisi, sebagai contoh dapat dilihat pada Gambar 2.

dengan *PoS Tagging* berbasis aturan (*rule based*) walaupun memiliki tingkat keakuratan yang tinggi 100% dalam korpus, dan *PoS HMM* sebesar 99,29%, penelitian ini belum memberikan gambaran solusi tagging dalam kalimat yang mengacu kepada jenis verba kopulatif, eksistensial dan ekuatif dalam bahasa Indonesia.

Pramudita et al (2016) dengan riset *part of speech tagging* berbasis aturan dan distribusi probabilitas maksimum *entropy* dengan objek kajian bahasa Jawa Krama. Data dalam penelitian ini diambil korpus kamus buku kosakata bahasa Jawa dengan tagset kategori leksikal N dengan penn Tree Tags NN, V, Adj, Adv, Knj, Prp, Kh, So, Em, Pr Sym, dan puntuasi yang menghasilkan maksimum *entropy* dan *rule based* dengan akurasi tertinggi 97,75% untuk data yang berbeda dari data training. Penggunaan Rule Based hanya bertumpu pada kategori leksikal saja walaupun tingkat keakuratannya sangat tinggi.

Syahroni (2019) mengembangkan aplikasi penentuan kategori dan fungsi sintaksis kalimat bahasa Indonesia dengan pemanfaatan Natural Language Toklit (NLTK) dan memodifikasinya menjadi 6 level frasa NP (Noun Phrase), VP (Verb Phrase), PP (Preposition Phrase), AP (Adjective Phrase), ADVP (Adverb Phrase) dan NUMP (Numerical Phrase) dengan akurasi 93,2 %, dan pada fungsi sintaksis dari 6 level frasa terhadap kalimat bahasa Indonesia diubah menjadi Subjek, Predikat, Objek, Keterangan dan Pelengkap untuk penentuan klausa diperoleh akurasi sebesar 91,12%. Penelitian ini belum lengkap karena tidak memberikan gambaran terhadap peran sintaksis A (agen) dan (P).

Dewi & Ubaidi (2020) melakukan penelitian *PoS Tagging* bahasa Madura dengan metode Brill Tagger dengan latar belakang penggunaan bahasa Madura dengan perbaikan dan penambahan data korpus dengan peningkatan akurasi leksem-leksem dari 87,43% menjadi 93,16%. Jika tidak memperhatikan *out of vocabulary* akurasi mencapai di atas 80% dengan akurasi tertinggi 86,67%. *Lexical rule* (aturan kategori) dengan pelabelan kata menjadi fokus penelitian ini, sehingga pada tataran fungsi dan peran sintaksis belum dianalisis secara lengkap. Nilai akurasi yang didapat dengan menggunakan metode Brill Tagger hanya muncul pada tataran kategori leksikal Verba Transitif (VBT), Verba Intransitif (VBI), Adjective (JJ), Adverb (RB), Common Noun, Proper Common Noun, Genitive Common Noun, Personal Pronoun, Locative Pronoun, Primary Numeral, Collective Numeral, Distributive Numeral, Irregular Numeral, Kata Bantu Bilangan, WH Pronoun, WH Adverb, Determiner, Article, Preposition, Coordinate Conjunction, Subordinate Conjunction, Particle, Interjection, Positive Modal, Negative Modal, Symbol (Sym), Sentence Terminator (ST), Comma (,), Elipsis (...), Colon (:), Semi Colon (;), Open Pharantheis, Close Pharanthesis (OP), Close Pharanthesis (CP), Quotation (QT), Dash (DASH), dan Slash (GM) dengan peningkatan jumlah korpus hanya pada tataran leksikal saja.

POS tagging bahasa Indonesia maupun bahasa daerah telah banyak diteliti menggunakan *PoS Tagging* dengan berbagai metode dan pendekatan baik yang berbasis aturan (*rule based*) maupun kemungkinan (*probabilitas*). Pemanfaatan *PoS Tagging* dengan memanfaatkan NLP (*Natural Language Processing*) telah banyak menghasilkan Tag yang memuaskan dengan tingkat akurasi yang tinggi. Hal ini terlihat dari penelitian yang *PoS Tagging* bahasa oleh Gunawan et al (2019) dengan membuat *rule* huruf kapital dalam bahasa Indonesia. Yazid & Fatwanto (2018) membuat tag kategori leksikal secara umum dalam bahasa Indonesia untuk mengurai kata ambigu dengan metode *rule-based*. Begitu juga dengan penelitian terkait fungsi Pramudita et al (2016) pada bahasa Jawa Krama yang sudah menyentuh level fungsi.

Analisis kesenjangan dalam penelitian ini adalah belum adanya penelitian lengkap terkait dengan *PoS Tagging* sintaksis bahasa Indonesia yang lengkap membahas penuh pada semua bentuk kategori (leksikal) dengan adanya predikat selain verba (kopulatif, ekuatif, dan eksistensial), fungsi, dan peran semantis meliputi agen dan pasien dalam membuat *PoS Tagging Rule* yang tepat dan utuh dalam tataran gramatika bahasa Indonesia.

Penelitian ini bertujuan untuk memperbaiki data korpus bahasa Indonesia terutama kuantitas dan kualitas pada tataran gramatikal dan dapat digunakan dalam pengklasifikasian kelas kata, fungsi kelas kata, dan peran semantis untuk mendukung riset bahasa dan pembelajaran bahasa terutama pengayaan klasifikasi leksikon dalam KBBI (Kamus Besar Bahasa Indonesia) secara elektronik karena dalam leksigrafi, faktor yang terpenting adalah tahap pengumpulan data. Sehingga kelompok besar data dalam KBBI yang mencakup kata umum



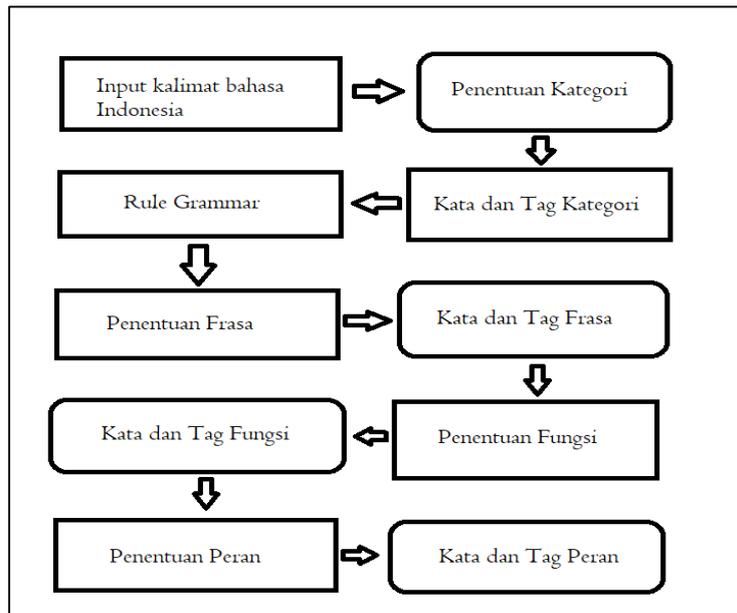
(*command word*) dan nama diri (*proper name*). Dalam hal *Command Word* yang meliputi empat aspek yaitu kelas kata, bentuk leksikal, struktur leksikal dan tipe leksikal dapat terpenuhi secara sempurna dengan adanya *PoS Tagging* fungsi, kategori dan peran sintaksis bahasa Indonesia.

Berdasarkan hal tersebut, tujuan dalam penelitian ini adalah berfokus pada algoritma penentuan kelas kata atau kategori sintaksis (leksikal), *rule-based* fungsi sintaksis dan peran semantik. Sebagai bentuk alternatif dan solusi dengan berkembangnya teknologi komputasi, kajian gramatikal bahasa Indonesia dapat dijadikan bahan pengembangan penelitian dengan memanfaatkan komputasi dan sains data. Hal tersebut menjadikan penelitian sintaksis bahasa Indonesia sebagai unsur strategis dalam memberikan kontribusi terhadap pemahaman tata bahasa secara umum dan lebih khusus pada hal-hal yang masih dianggap ambigu dan membingungkan terutama dari segi gramatikal fungsional, kategorial, dan semantik. Penelitian ini juga memberikan kontribusi kamus KBBI (Kamus Besar Bahasa Indonesia) elektronik yang memuat kriteria formal (mikrostruktur) terutama pada sublema kategori, fungsi, dan peran sintaksis yang belum dijangkau terutama peran sintaksis (semantis) yang memuat agen dan pasien dalam bahasa Indonesia.

METODE

Data dalam penelitian ini adalah kualitatif yang berupa bahan korpus (Carroll, 2013). Analisis data kualitatif terhadap bahan korpus menggunakan teknik berbasis probabilitas (*probability-based*) untuk menelaah tataran gramatikal leksikon dalam kalimat (Jurafsky & Martin, 2000; Widhiyanti & Harjoko, 2013; Pramudita et al., 2016). Sementara untuk pendekatan berbasis aturan (*Rule-Based*) digunakan untuk mengonstruksi kata dengan model *statistic scholastic tagger* untuk memperkirakan nilai probabilitas data pada penentuan. Data dalam penelitian ini menggunakan data training untuk penentuan kategori dan data uji. Adapun untuk fungsi sintaktis dan peran menggunakan data 990 kalimat yang telah ditentukan peran agen dan pasiennya. Data *training* berupa berkas korpus *The corpus ind_mixed_2013 Leipzig Corpora* yang merupakan sekumpulan jumlah kalimat berbasis korpus. *Raw data* dalam berkas *The corpus ind_mixed_2013 Leipzig Corpora* memuat berkas daring berita pada laman (Effendi & Muchammadun, 2018; Rajeg et al., 2019; Dewi & Ubaidi, 2020; Yusuf & Puspita, 2020; Yuki, 2021).

Langkah yang dilakukan untuk pengumpulan data yaitu (1) *preprocessing* dengan melakukan filter terhadap kalimat input agar dapat diproses oleh *POS Tagging*, (2) memanfaatkan *PoS Tagging* korpus *The corpus ind_mixed_2013 Leipzig Corpora* untuk memberikan tag pada setiap kalimat input, (3) mengubah level kata hasil *PoS Tagger* sesuai dengan kategori sintaksis FN (frasa nomina), FAdj (frasa adjektiva), FV (frasa verba) terkait verba dalam penelitian ini akan ditambahkan idenfikasi verba kopulatif, verba, ekuatif dan eksistensial, F Prep (frasa preposisi), Fadv (frasa Adverbia), (4) menentukan kedudukan fungsi sintaksis S (subjek), P (predikat), O (objek), K (keterangan), dan P (pelengkap) (5) menentukan kedudukan peran semantik A (agen) dan P (pasien). Lebih lanjut langkah dalam pengumpulan data dalam penelitian ini dapat dilihat pada Gambar 3.



Gambar 3

Flow Chart Alur Penelitian PoS Tagging Kategori, Fungsi dan Peran dan Rule-Based dan Probability

Model tersebut di-*training* dan di-*testing* dengan data korpus bahasa Indonesia Leipzig dari library Flair NLP. Setelah terbentuk model prediksi kategori tersebut, maka dilakukan penerapan prediksi kategori pada 990 kalimat baru dari korpus *The corpus ind_mixed_2013 Leipzig Corpora*. Hasil prediksi kategori pada sejumlah kalimat tersebut selanjutnya dikombinasikan dengan *rule* grammar frasa dan menjadi input untuk penentuan frasa. Penentuan frasa dilakukan dengan menyusun algoritma fungsi dengan *output* berupa frasa dan tak frasa. Kemudian, *output* tersebut menjadi input dari algoritma penentuan fungsi dengan *output* berupa kata dan tak fungsi. Terakhir, *output* kata dan tak fungsi menjadi input dari algoritma penentuan peran dan menghasilkan *output* berupa kata dan tak peran. Hasil akhir dari penelitian ini adalah 990 kalimat baru dari corpus Leipzig yang telah ditentukan kategori, fungsi, dan perannya untuk masing-masing kata.

HASIL DAN PEMBAHASAN

Hasil *Tagging* berbasis aturan *Rule-Based* pada data korpus *The corpus ind_mixed_2013 Leipzig Corpora* menunjukkan hasil akurasi setelah diberikan tambahan *Tag* pada kategori sintaksis terutama verba secara umum F-Score (mikro) 0.9253, F-Score (makro) 0.8576 dengan akurasi 0.9253. Secara lebih khusus pada penelitian ini, hasil tag Verba (VERB) memiliki rasio prediksi yang positif setelah adanya penambahan bagian verba kopulatif (adalah, ialah dan \emptyset), ekuatif (sama, seperti, Adj +*se*-*ekuatif*, dan N + *se*-*unit*), dan eksistensial (ada) dengan rasio prediksi benal (recall: 0.9256). Hal tersebut juga berpengaruh pada *Tag* dengan *rule-based* fungsi sintaksis dengan verba kopulatif, ekuatif, dan eksistensial yang menduduki fungsi predikat pada konstruksi sintaktis. Selanjutnya, pada peran 57 semantik yang merupakan temuan baru dalam penelitian ini dengan rasio prediksi *true* 82.58% pada Agen (A) dan 9.68% pada pasien (P). Tingkat akurasi ini dapat dilihat pada Tabel 3.



Tabel 3
 Tingkat Akurasi Kategori dengan Penambahan *Rule-Based* Verba Kopulatif, Ekuatif, dan Eksistensial

No	By Class	Precision	Recall	f1-score	support
1	NOUN	0.8789	0.9076	0.8930	2488
2	PROPN	0.9155	0.9280	0.9217	2207
3	PUNCT	0.9970	0.9994	0.9982	1645
4	VERB (ada)	0.9562	0.9256	0.9407	1250
5	ADP	0.9682	0.9506	0.9593	1153
6	ADV	0.8061	0.8624	0.8333	487
7	PRON	0.9791	0.9398	0.9590	498
8	NUM	0.9269	0.9656	0.9458	407
9	ADJ	0.8219	0.7109	0.7624	422
10	DET	0.9449	0.9220	0.9333	372
11	CCONJ	0.9799	0.9093	0.9433	375
12	SCONJ	0.7756	0.8521	0.8121	142
13	AUX	0.9897	1.0000	0.9948	96
14	PART	0.9149	0.8776	0.8958	49
15	SYM	1.0000	0.8667	0.9286	30
16	X	0.0000	0.0000	0.0000	1

Pada Tabel 3 dapat disimpulkan bahwa, *Accuracy* adalah rasio prediksi benar dengan keseluruhan data. *Precision* adalah rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. *Recall* adalah rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. *F1-score* merupakan perbandingan rata-rata *precision* dan *recall* yang dibobotkan.

Pola pada penelitian Syahroni (2019) dengan *rule regexpParser* dengan NLTK (*Natural Language Toklit*) dengan pencocokan bahasa Indonesia dengan mengekstrak 6 level frasa yaitu FN (Frasa Nomina), FV (Frasa Verba), Frasa Preposisi (FP), FAdj (Frasa Adjektiva), FAdv (Frasa Adverbia) dan FN (Frasa Numeralia). Penelitian ini belum menunjukkan hasil yang spesifik dan mendalam terkait dengan bagian FV (Frasa Verba) dengan adangan verba kopulatif, ekuatif dan eksistensial walaupun tingkat akurasi 93,32%. Pembentukan aturan (*rule-based*) dengan menambahkan partikel verba sebagai fenomena sintaktis dapat menjadi pedoman dalam menentukan bagian yang perlu diperhatikan dalam fenomena verba sebagai inti (*core*) (Jia et al., 2020).

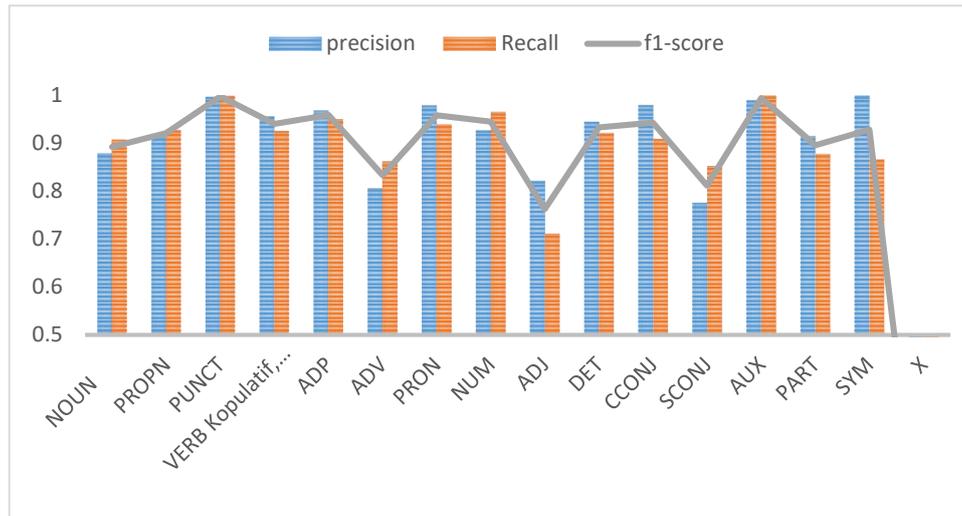
POS Tagging Berbasis Probability Kategori Sintaksis

Berdasarkan hasil *Tag* dengan *probability-based* kategori, khususnya predikat dalam bahasa Indonesia dapat diduduki oleh verba kopulatif, eksistensial dan ekuatif. Selanjutnya, data korpus kategori dengan *probability-based* dapat dilihat pada Gambar 4.

```
def kategori_f(x):
    for I in range(len(x)):
        sentence = Sentence(x['KALIMAT_k'].iloc[i])
        tag_pos.predict(sentence)
        x.loc[I,'Kategori']=sentence.to_tagged_string()
```

Gambar 4. Syntax Penentuan Kategori

Fungsi di atas mengolah input berupa tabel, tabel tersebut memiliki tiga kolom, yaitu kolom KALIMAT_k, kolom AGEN, dan kolom PASIEN, serta 990 baris kalimat. Untuk setiap baris, diterapkan fungsi prediksi tak kategori pada setiap kata yang terkandung di tiap kalimat. *Output* dari fungsi ini adalah berupa tabel input dengan tambahan satu kolom yaitu kolom KATEGORI. Kolom kategori tersebut berisi masing-masing kata beserta tak kategorinya. Hasil dari fungsi syntax kategori ini dapat digambarkan pada Gambar 5.



Gambar 5. Diagram Akurasi Kategori Sintaksis dengan *PoS Tagging Probability-Based*

Diagram akurasi ini didapatkan dari hasil *Tag* semua kategori yang ada pada data korpus *The corpus ind_mixed_2013 Leipzig Corpora*. Hasil dari syntax penentuan frasa data korpus tersebut dapat dilihat pada Gambar 6.

```
def frasa_f(x):
    for I in range(len(x)):
        taggedI = x['Kategori'].iloc[i]
        taggedI = str(taggedI)
        taggedI = taggedI.replace("<", "")
        taggedI = taggedI.replace(">", "")
        taggedI = Convert(taggedI)
        taggedI = [taggedI[y:y+2] for y in range(0, len(taggedI), 2)]
        taggedI = [tuple(l) for l in taggedI]
        tree = chunkParser.parse(taggedI)
        tree = str(tree)
        x.loc[I, 'Frasa'] = tree
```

Gambar 6. Syntax Penentuan Frasa

Fungsi di atas mengolah input berupa tabel, tabel tersebut memiliki empat kolom, yaitu kolom KALIMAT_k, kolom AGEN, kolom PASIEN, kolom KATEGORI, dengan 990 baris. Untuk setiap baris, diterapkan fungsi prediksi tak frasa pada setiap frasa yang terkandung di tiap kalimat. *Output* dari fungsi ini adalah berupa tabel input dengan tambahan satu kolom yaitu kolom FRASA. Kolom tersebut berisi masing-masing frasa beserta tak frasanya.

PoS Tagging Berbasis Rule-Based Fungsi Sintaksis

Secara gramatikal, kategori sintaksis dapat menduduki fungsi sintaksis dalam sebuah klausa atau kalimat. Tahapan *rule-based* berikutnya adalah membuat *rule* dengan ketentuan seperti yang dikemukakan oleh (Kridalaksana, 1988; Alwi, 2003) pada unsur inti kalimat atau klausa yaitu S dan P. Ciri Subjek yang di *Tag* pada data korpus *The corpus ind_mixed_2013 Leipzig Corpora* adalah S yang dapat berupa NOM, Frasa Nomina, atau klausa. Pada unsur predikat dengan *rule* grammar pengisi unsur predikat adalah VERB (kopulatif, ekuatif dan eksistensial), VERB PHRASE, ADJ, ADJ PHRASE, dan NOUN



PHRASE. Hasil syntax penentuan fungsi kategori (leksikal) yang mengisi fungsi sintaksis dapat dilihat pada Gambar 7.

```
def spok_f(x):
    for I in range(len(x)):
        tree = x['Frasa'].iloc[i]
        tree = nltk.tree.Tree.fromstring(str(tree))
        ptree = ParentedTree.convert(tree)
        for subtree in ptree.subtrees():
            if subtree.label() == 'VP':
                if subtree.left_sibling() is not None:
                    if type(subtree.left_sibling()) is str:
                        x.loc[I,'Subjek'] = str(subtree.left_sibling())
                        x.loc[I,'Predikat'] = str(subtree.leaves()[0])
                        x.loc[I,'Objek'] = str(subtree.leaves()[1:])
                    else:
                        if subtree.left_sibling().label() == 'NP':
                            x.loc[I,'Subjek'] = str(subtree.left_sibling())
                        elif subtree.left_sibling().label() == 'NUMP':
                            x.loc[I,'Subjek'] = str(subtree.left_sibling())
                        elif subtree.left_sibling().label() == 'PP':
                            x.loc[I,'Keterangan'] = str(subtree.left_sibling())
                        else:
                            x.loc[I,'Predikat'] = str(subtree.leaves())
                if subtree.label() == 'VP':
                    if subtree.right_sibling() is not None:
                        if type(subtree.right_sibling()) is str:
                            x.loc[I,'Predikat'] = str(subtree.leaves()[0])
                            x.loc[I,'Objek'] = str(subtree.leaves()[1:])
                            x.loc[I,'Keterangan'] = str(subtree.right_sibling())
                        else:
                            if subtree.right_sibling().label() == 'NP':
                                x.loc[I,'Objek'] = str(subtree.right_sibling())
                            elif subtree.right_sibling().label() == 'NUMP':
                                x.loc[I,'Objek'] = str(subtree.right_sibling())
                            elif subtree.right_sibling().label() == 'PP':
                                x.loc[I,'Keterangan'] = str(subtree.right_sibling())
                            else:
                                x.loc[I,'Predikat'] = str(subtree.leaves())
                    if ((subtree.right_sibling() is None) and (subtree.left_sibling() is None)):
                        x.loc[I,'Subjek'] = str(subtree.leaves()[0])
                        x.loc[I,'Predikat'] = str(subtree.leaves()[1])
```

Gambar 7. Syntax Penentuan Fungsi

Fungsi di atas mengolah input berupa tabel, tabel tersebut memiliki lima kolom, yaitu kolom KALIMAT k, kolom AGEN, kolom PASIEN, kolom KATEGORI, dan kolom FRASA, dengan 990 baris. Untuk setiap baris, diterapkan prediksi tag fungsi. *Output* dari ini adalah berupa tabel input dengan

tambahan empat kolom yaitu kolom SUBJEK, PREDIKAT, OBJEK, dan KETERANGAN. Kolom tersebut berisi kata yang menduduki fungsi tertentu dalam sebuah kalimat.

Apabila verbanya adalah tindakan (*action verb*) dan subjek dan objeknya adalah manusia dan animate maka objeknya adalah pasien dan subjeknya adalah agen. Selanjutnya, apabila objeknya bukan manusia dan animate maka perannya adalah sebagai tema. Hal ini seperti yang kemukakan oleh Tallerman (2014) dan Rober (2001) jika verbanya transitif subjeknya adalah agen dan objeknya pasien.

Rule-Based Peran Sintaksis Agen

Hasil *tag* dari kedudukan partisipan sesuai dengan perannya pada posisi agen muncul pada subjek makhluk hidup yang melakukan aksi dengan kategori N dan PRON. Hal ini memiliki relevansi dengan apa yang dikemukakan oleh Carina (2018) bahwa agen/pelaku adalah makhluk yang secara sadar melakukan tindakan yang dinyatakan oleh verba. Agen pada konteks *Tag Rule-Based* ini berada dalam aksi proses yang selalu menduduki fungsi subjek pada kalimat aktif dan akan menjadi peran pasien pada kalimat pasif dengan hasil prediksi 82,58%. Hasil prediksi akurasi kemunculan agen pada data korpus *Leipzig Indonesian Mix 2013* dapat dilihat pada Tabel 4 dan Gambar 8 syntax penentuan agen berikut.

Tabel 4
Rule-Based Peran Sintaksis Agen Data Korpus Leipzig Indonesian Mix 2013

Agen Tag	Row Label	Agen_eq (prediksi) rule-based		Jenis kalimat
		TRUE	FALSE	
Otoritas local (blank)	NP	-	√	Nominatif akusatif (Aktif)
	NP/NP/PRON	-	√	
		√	-	Nominatif akusatif (Aktif)
tersangka	N	√	-	Nominatif akusatif (Aktif)
KPU	N	√	-	Nominatif akusatif (Aktif)
wahyu	PRON	√	-	Nominatif akusatif (Aktif)
wartawan	N	√	-	Nominatif akusatif (Aktif)
subono	PRON	√	-	Nominatif akusatif (Aktif)
suaranya	PRON	√	-	Nominatif akusatif (Aktif)
solksjaer	N	√	-	Nominatif akusatif (Aktif)
sitorus	PRON	√	-	Nominatif akusatif (Aktif)
seniman	PRON	√	-	Nominatif akusatif (Aktif)
saya	N	√	-	Nominatif akusatif (Aktif)
sahabatnya	PRON	√	-	Nominatif akusatif (Aktif)
rozaq	PRON	√	-	Nominatif akusatif (Aktif)
presiden	PRON	√	-	Nominatif akusatif (Aktif)
pelaku	N	√	-	Nominatif akusatif (Aktif)

```
data['Prediksi_Agen'] = pd.np.where(data['Predikat'].str.contains('me'), data['Subjek'], pd.np.where(data['Predikat'].str.contains('di'), data['Objek'], ''))
```

Gambar 8. Syntax Penentuan Agen

Rule-Based Peran Sintaksis Pasien

Peran sintaksis pada jenis pasien setelah dibuat *rule-based* dengan *PoS Tagging* dengan hasil prediksi 9.68%, ditemukan beberapa kategori *rule* dengan prediksi N dan PRON yang berperan sebagai pasien pada kalimat pasif dengan fungsi sebagai objek dan subjek pada kalimat yang anti pasif dengan tipe ergative. Hal ini berhubungan dengan penentuan argumen dalam bahasa yang melekat pada sisi transitivitas verbanya. Dixon (1994) mengemukakan bahwa jika argumennya adalah agen maka posisinya adalah subjek dan pasien adalah subjek secara fungsional. Hal ini juga berkorelasi dengan apa yang dikemukakan oleh Jufrizal (2008) dan Artawa (2020) bahwa dalam pola Pasif dan ergatif sama-sama



melibatkan paling tidak properti S sama dengan P daripada A dan tidak dapat dipasifkan dengan adanya kendala morfologi tertentu. Hasil prediksi *rule-based* pada peran sintaksis pasien ini dapat dilihat pada Tabel 5 dan Gambar 9.

Tabel 5
 Hasil *Rule-Based* Peran Sintaksis Pasien Data Korpus *Leipzig Indonesian Mix 2013*

Agen Tag	Row Label	Pasien_eq (prediksi) rule-based		Jenis kalimat
		TRUE	FALSE	
acara doa	NP	-	√	Pasif/Ergatif (Anti-Pasif)
Indonesia	N	-	√	Pasif/Ergatif (Anti-Pasif)
instansi daerah	NP	-	√	Pasif/Ergatif (Anti-Pasif)
isolasi	N	-	√	Pasif/Ergatif (Anti-Pasif)
jajarannya	NP	-	√	Pasif/Ergatif (Anti-Pasif)
Jakarta	N	-	√	Pasif/Ergatif (Anti-Pasif)
jambu	N	-	√	Pasif/Ergatif (Anti-Pasif)
kantor kecamatan	NP	-	√	Pasif/Ergatif (Anti-Pasif)
komentar warganet	NP	-	√	Pasif/Ergatif (Anti-Pasif)
kunjungan	V	-	√	Pasif/Ergatif (Anti-Pasif)
manchester city	NP	-	√	Pasif/Ergatif (Anti-Pasif)
nurhadi resky	PRON	√	-	Pasif/Ergatif (Anti-Pasif)
nyoman giri prasta	PRON	√	-	Pasif/Ergatif (Anti-Pasif)
oknum dosen	NP	√	-	Pasif/Ergatif (Anti-Pasif)
orang-orang	N	√	-	Pasif/Ergatif (Anti-Pasif)
para demonstran	NP	√	-	Pasif/Ergatif (Anti-Pasif)
para pekerja seni	NP	√	-	Pasif/Ergatif (Anti-Pasif)
para siswa	NP	√	-	Pasif/Ergatif (Anti-Pasif)
pemain	N	√	-	Pasif/Ergatif (Anti-Pasif)
pemerintah libia	NP	√	-	Pasif/Ergatif (Anti-Pasif)
pencuri	N	√	-	Pasif/Ergatif (Anti-Pasif)

```
data['Prediksi_Pasien'] = pd.np.where(data['Predikat'].str.contains('me'), data['Objek'], pd.np.where(data['Predikat'].str.contains('di'), data['Subjek'], ''))
```

Gambar 9. Syntax Penentuan Pasien

Telaah kategori, fungsi dan peran semantik dalam bahasa Indonesia sudah banyak diteliti. Pendekatan linguistik komputasi maupun linguistik korpus juga banyak yang membahas fungsi dan kategori. Keunggulan dari hasil penelitian ini adanya rule baru yang dibuat berdasarkan data korpus *The corpus ind mixed 2013 Leipzig Corpora* terbaru terkait aturan *PoS Tagging* dengan subjek intransitif bersamaan dengan objek transitif (absolutif) dan berbeda dengan subjek transitif (ergatif). Hal ini sejalan dengan yang dikemukakan oleh [Coon & Abenina-Adar \(2013\)](#) bahwa potensi kemunculan ergatif dalam bahasa akan ada dengan adanya gejala morfologis.

Kecenderungan secara umum penelitian fungsi, kategori, dan peran sintaksis bahasa Indonesia adalah dalam bentuk nominatif akusatif yang menurunkan diatesis aktif dan pasif sedangkan ergatif belum banyak disentuh. Hal tersebut menjadi menarik jika melihat pola morfosintaktis dalam bahasa Indonesia yang sangat mungkin dibuat rule ergatif dengan berbagai sudut pandang linguistik bahwa fenomena ergatif dalam bahasa adalah fenomena yang unik ([Antonio, 2017; Artawa, 2020; Kridalaksana, 1988; Mughni 2021; Pye, 1990; Satriya & Wisnu, 2016; Sherly, 2019](#)).

SIMPULAN

Tingkat akurasi *PoS Tagging probability-based* pada data training kategori sintaksis (leksikal) setelah diberikan penambahan verba kopulatif, ekuatif, dan eksistensi menunjukkan hasil yang memuaskan yaitu 92,53%. Hasil tersebut digunakan untuk menentukan *rule-based* fungsi sintaksis terutama predikat

yang diduduki selain verba dalam bahasa Indonesia. Pada peran sintaksis, penentuan kedudukan peran agen dan pasien dapat ditentukan dari hasil penentuan kategori dan fungsi dengan hasil yang prediksi hasil *rule-based* yang didapatkan untuk agen 82,58% dengan prediksi pola N dan PRON pada kalimat aktif sebagai agen dan pasif sebagai pasien pada tipe nominative akusatif catatan verbanya transitif ada objek, jika verbanya intransitif maka tidak ada objek. Jika verbanya kopulatif maka konstruksinya terdiri dari Subjek+Predikat+Komplemen, ekuatif dengan konstruksi S+P yang predikatnya itu diduduki oleh nomina, adjektiva, numeralia dan frasa preposisi dengan catatan *rule-based* yang digunakan adalah agen yang berjenis animate. Sedangkan untuk pasien dengan prediksi *rule-based* 9,68% menunjukkan adanya konstruksi pasif, ergative, dan anti-pasif dengan subjek adalah pasien dan bukan objek. Dengan adanya *rule based* fungsi, kategori, dan peran sintaksis bahasa Indonesia pada penelitian sangat berimplikasi kepada pengajaran dan telaah bahasa Indonesia agar memperhatikan sudut pandang lain dalam masalah penentuan verba secara kategori (leksikal), subjek dan objek secara fungsional hingga pasien dalam tinjauan peran semantis. Tata bahasa Indonesia dengan *PoS Tagging* dengan berbagai pendekatan korpus dan komputasi merupakan hal yang penting untuk dikaji menghadapi era digitalisasi pada masa sekarang. Kajian linguistik mikro dengan memanfaatkan data korpus diharapkan mampu memperbaiki telaah yang belum lengkap seperti adanya fenomena kemunculan verba kopulatif, ekuatif, dan eksistensial hingga tipologi bahasa Indonesia yang unik dengan kemunculan ergatif sehingga hasil dari *PoS Tagging* ini dapat dijadikan data aturan untuk membuat aplikasi tata bahasa Indonesia yang lebih mapan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Pendidikan Bahasa dan Sastra Indonesia, Universitas Tidar yang telah memberikan kepercayaan untuk melaksanakan penelitian penugasan kepakaran program studi untuk mendukung proposal Simlitabmas. Ucapan terima kasih juga penulis sampaikan kepada Ibu Asri Wijayanti, S.Pd., M.A selaku Koordinator Program Studi Pendidikan Bahasa dan Sastra Indonesia, Bapak Isyadi Shalima, S. S., M.A selaku anggota penelitian, Ratih Ngestrini, S. ST., M. Sc dan Anandika Dian Pramesti, S.Tr. Stat yang telah membantu dalam mengolah data, Bapak Dr. Muhammad Yusdi, M.Hum yang telah memberikan masukan dan saran untuk artikel ini serta Muhammad Abdul Hanif dan Deri Nasrullah mahasiswa Pendidikan Bahasa dan Sastra Indonesia yang telah membantu mengumpulkan data penelitian.

DAFTAR PUSTAKA

- Alwi, H. et al. (2003). *Tata bahasa baku bahasa Indonesia: Edisi Ketiga*. Jakarta: Balai Pustaka.
- Ambiya, M. Z. (2018). Keberlakuan nomina sebagai predikat dalam kalimat bahasa Indonesia: Kajian sintaksis (Noun Existence as predicate in sentence of Indonesian language: A study on syntax). *Ranah: Jurnal Kajian Bahasa*, 7(1), 49-68. <https://doi.org/10.26499/rmh.v7i1.543>
- Antonio, S. (2017). *A tentative corpus-based approach to English ergative verbs*. July. <https://doi.org/10.13140/RG.2.2.12120.19208>
- Arka, I. W. (2018). Lexical-functional grammar: Some of its basic principles and its challenge in its application to the languages of Indonesia. *Pertemuan Linguistik Pusat Kajian Bahasa dan Budaya Atma Jaya 16*. Jakarta: Yayasan Obor Indonesia.
- Artawa, K. (2020). Pemarkahan diatesis bahasa Indonesia. *Mozaik Humaniora*, 20(1), 26-?. <https://doi.org/10.20473/mozaik.v20i1.15128>
- Carina Silberer, M. P. (2018). Grounding semantic roles in images. In B. Brussels (Ed.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 2616–2626). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D18-I282>
- Carroll, J. (2013). *The Routledge handbook of corpus linguistics*. London: Routledge.
- Cholisi, F. (2013). A lexical-functional grammar representation. *Jurnal Sosial Humaniora*. 6(1), 46-61. <https://doi.org/10.12962/j24433527.v6i1.610>
- Comrie, B. (1981). *Language universals and linguistics typology*. New Jersey: Basil Blackwell.



- Coon, J., & Abenina-Adar, M. (2013). *Ergativity in linguistics*. Oxford: Oxford University Press.
- Dewi, N. P., & Ubaidi, U. (2020). *PoS tagging bahasa Madura dengan menggunakan algoritma Brill Tagger*. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(6), 1121-1128. <https://doi.org/10.25126/jtiik.2020722449>
- Dixon, R. M. (1994). *Ergativity*. Cambridge: Cambridge University Press.
- Effendi, D. E., & Muchammadun. (2018). "Happiness" in bahasa Indonesia and its implication to health and community well-being. *Asian EFL Journal*, 20(8), 279-291.
- Falk, Y. (2011). *Lexical-functional grammar*. Oxford: Oxford University Press.
- Gunawan, D., Amalia, A., & Maringga, O. N. (2019). Building the Application to identify incorrect capital letters writing in bahasa Indonesia. *Journal of Physics: Conference Series*, 1235(1), 1-7. <https://doi.org/10.1088/1742-6596/1235/1/012108>
- Jia, J. K., Shao, Y. bin, Long, H., & Du, Q. Z. (2020). A natural language sentence analysis algorithm based on word order modifier syntax rules. *Procedia Computer Science*, 166, 496-500. <https://doi.org/10.1016/j.procs.2020.02.058>
- Jufrizal, J. (2008). Tipologi linguistik: Dasar kerangka teori dan arah kajiannya. *Linguistika*, 1(3), 23-35. Retrieved from <http://jurnalvivid.fib.unand.ac.id/index.php/lingkul/index>
- Jurafsky, D., Martin, J. H. (2000). *Speech and language processing*. Britain: Prentice Hall.
- Kridalaksana, H. (1988). *Beberapa prinsip perpaduan leksem dalam bahasa Indonesia*. Jakarta: Kanisius.
- Kridalaksana, H. (2008). *Kamus linguistik edisi keempat*. Jakarta: Gramedia Pustaka Utama.
- Mallinson, G. dan B. J. B. (1981). *Language typology: Cross-linguistic studies in syntax*. North-Holland Publishing Company.
- Mughni, M. B. M. Y. (2021). Perbandingan kata kerja tak akusatif dengan kata kerja tak ergatif bahasa Melayu: Analisis Aktionsart. *Jurnal Linguistik*, 24(1). 19-32. doi: 10.31604/linguistik.v3i1.19-38
- Musgrave, S. (2013). Functional categories in the syntax and semantics of Malay. *NUSA: Linguistic Studies of Languages in and around Indonesia*, 55, 135-152. Retrieved from <http://repository.tufs.ac.jp/bitstream/10108/74330/2/nusa5509.pdf>
- Pramudita, H. R., Utami, E., & Amborowati, A. (2016). Pengaruh part of *speech tagging* berbasis aturan dan distribusi probabilitas maximum entropy untuk bahasa Jawa Krama. *Jurnal Buana Informatika*, 7(4), 10-25. <https://doi.org/10.24002/jbi.v7i4.764>
- Pye, C. (1990). The acquisition of ergative languages. *Linguistics*, 28(6), 1291-1330. <https://doi.org/10.1515/ling.1990.28.6.1291>
- Rajeg, G. P. W., Denistia, K., & Rajeg, I. M. (2019). Working with a linguistic corpus using r: An introductory note with Indonesian negating construction. *Linguistik Indonesia*, 36(1), 1-36. <https://doi.org/10.26499/li.v36i1.71>
- Ramlan. (1987). *Ilmu bahasa Indonesia, sintaksis*. Yogyakarta: CV Karyono.
- Rober, D. Van Valin, Jr. R. J. L. (2001). *Syntax: Structure, meaning and Fuction*. Cambridge: Cambridge University Press.
- Satriya, S., & Wisnu, T. (2016). Pasif perifer: Struktur ergatif dalam bahasa Indonesia. *Metalingua*. 14(2), 233-242. <http://dx.doi.org/10.26499/metalingua.v14i2.199>
- Sherly, Novita, M. (2019). *Pembentukan verba ergatif dalam bahasa Hokkien: Kajian morfosintaksis*. 26(1), 8-17. <https://doi.org/https://doi.org/10.24843/ling.2019.v26.i01.p02>.
- Sneddon, J. N., Adelaar, A., Djenar, D. N., & Ewing, M. C. (2010). *Indonesian Reference Grammar*. Oxford: Allen & Unwin.
- Song, J. J. (2014). *Linguistic typology* (1st Edition). Britain: Oxford University Press.
- Syahroni, Abd. Wahab, H. (2019). Aplikasi penentuan kategori dan fungsi sintaksis kalimat bahasa Indonesia. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, 1(3), 13-20 <https://doi.org/https://doi.org/10.30743/infotekjar.v4i1.1537>
- Tallerman, M. (2014). *Understanding syntax. (Fourth Edition)*. New York: Routledge.



-
- Tallerman, M. (2015). *Understanding syntax* (Fourth edi). New York: Routledge.
- Verhaar, J. W., & Alip, B. (1996). *Asas-asas linguistik umum*. Yogyakarta: Gadjah Mada University Press.
- Widhiyanti, K., & Harjoko, A. (2013a). *PoS tagging* bahasa Indonesia dengan HMM dan *rule based*. *Jurnal Informatika*, 8(2), 151-167. <https://doi.org/10.21460/inf.2012.82.125>
- Yazid, A. S., & Fatwanto, A. (2018). Penentuan kelas kata pada part of speech tagging kata ambigu bahasa Indonesia. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 157-166. <https://doi.org/10.14421/jiska.2018.23-05>
- Yuki, L. K. (2021). Study of the meaning of lelaki, bujang and bujangan words in the utilisation of Indonesian corpus. *Jurnal Sasindo UNPAM*, 9(1), 81-93. <https://doi.org/10.32493/sasindo.v9i1.81-93>
- Yusuf, K., & Puspita, D. (2020). Diachronic corpora as a tool for tracing etymological information of Indonesian-Malay Lexicon. *Register Journal*, 13(1), 153-182. <https://doi.org/10.18326/rgt.v13i1.153-182>