

Bare Demo of IEEEtran.cls for IEEECS Conferences

Abstract—The abstract goes here. DO NOT USE SPECIAL CHARACTERS, SYMBOLS, OR MATH IN YOUR TITLE OR ABSTRACT.

Keywords—component; formatting; style; styling;

```
=====
cite-citet-citep,      foto      ga      nongol,      refer-
ence  figure  di  dlm  text,  betulin  rumus.
=====
```

I. INTRODUCTION

There are several concepts that relate syntactic structures of a sentence with human's cognition, such as dependency. In a *dependency* theory, every linguistic unit is connected with each other by a link that represents direct asymmetrical syntactic relation forming an iterative structure attaching all the constituents within a sentence tesniere1959elements. This syntax framework contributes greatly to the discussion on how far a language development is shaped by the mechanical constraints of human cognition, particularly in relation to language acquisition and language application, as suggested by [?]. [?] proposed a linear distance that represents dependency between two words or constituents, which has been adopted into modern computational methods of quantifying dependency. This linear distance is measured by counting the constituents involved between the connected constituents. However, another important aspect in measuring dependency, in addition to the linear distance, is the position of the *head* in regards to its *dependent*, or the *head directionality*.

A. Measuring dependency

Words are the main constituents of syntax that carry with them the grammatical relation that connect one word with another tesniere1959elements. The term *dependency* itself is a derivative of a connection that refers to an asymmetrical relation between a superordinate and its subordinate (hudson1984word, heringer1993dependency, tesniere1959elements). tesniere1959elements, hudson1984word, liu2009chinese have all described in detail what a dependency structure between two constituents looks like. It can be summarized from these references that a dependency structure consists of a binary and asymmetrical relation between two linguistic units and the type of a dependency relation is usually indicated using a label on top of an arc linking the two constituents, as shown in 1. An arrow pointing from a more superior constituent to a more inferior one represents the hierarchical relation between

[width=0.8] pics/tautandependensi.png

Figure 1. Elements of a dependency relation

two constituents. The governor is also more commonly called as a *head* and the subordinate is generally called a *dependent*. In a more complex and recursive structure of dependency relations, such as a sentence, the sentence head is usually called a *root*.

The arc in 1 illustrates how a constituent must be kept active in working memory until both constituents are realized in an utterance, which result in the integration of meanings carried by both constituents hudson2003psychological, liu2008dependency. Therefore, the greater the dependency distance, the heavier the burden borne by the working memory. To measure a dependency distance (DD), a sentence is treated as a string of constituents $C_1...C_i...C_n$. Each constituent has a significant index that corresponds its position within the given sentence. For example, C_1 indicates that the constituent is located on the first position in the sentence. There are different approaches of measuring dependency distance. Largely, they involve the value of DD, which is obtained by subtracting directly linked constituents liu2008dependency, liu2017dependency, futrell2015large.

Previous substantial studies on measuring dependency, [?] and [?], used dependency length (DL) as their main approach. DL is the sum of absolute DD of an entire sentence, which can be defined as follows:

$$\sum_{i=1}^{n-1} |DD_i|$$

While other notable studies used a different approach called the mean dependency distance (MDD) liu2008dependency, liu2017dependency which divides DL with the number of dependency relations of an entire sentence. MDD can be defined as follows:

$$\frac{1}{n-1} \sum_{i=1}^{n-1} |DD_i|$$

B. Head directionality in Indonesian language

Hingga saat ini, belum ada konvensi mengenai direksionalitas induk terkait hierarki dan pengolahan dalam memori kerja. Namun, beberapa kajian telah dilakukan untuk memberikan penjelasan tambahan mengenai penerapan aturan tata bahasa. Beberapa penelitian pada awal berkembangnya teori dependensi, menemukan bahwa sebuah bahasa cenderung menerapkan arah relasi yang konsisten,

[width=0.5] pics/samebranching.png

Figure 2. percabangan-searah atau *same-branching*

[width=0.5] pics/mixedbranching.png

Figure 3. percabangan-beda-arrah atau *mixed-branching*

baik itu diawali-induk (*head-first* atau *head-initial*) maupun diakhiri-induk (*head-last* atau *head-final*)¹ (hawkins1994performance, radford1997syntactic, vennemann1994linguistic). Tindakan yang konsisten ini juga diasumsikan sebagai penerapan tata bahasa sebagai strategi untuk meminimalkan jarak antara induk dan konstituen terikatnya (hawkins1994performance, frazier1985syntactic). [?] juga telah berupaya untuk memberikan validitas berdasarkan data mengenai arah dependensi dan kaitannya dengan klasifikasi topologis sebuah bahasa.

Perbandingan percabangan pada 2 dan 3 mengilustrasikan tautan-tautan yang dibentuk oleh masing-masing tipe percabangan. Percabangan searah atau *same branching* dengan satu relasi yang berkelanjutan tidak mungkin akan secara konsisten terjadi, terutama pada kalimat yang lebih panjang sehingga asumsi ini dipertanyakan oleh [?] yang menganggap bahwa percabangan searah ini tidak selalu optimal dalam menggambarkan penampilan bahasa. [?] meneruskan asumsi ini berdasarkan penelitian [?] terhadap 625 bahasa yang mengungkapkan bahwa percabangan searah yang konvensional tidak dapat menggambarkan dengan baik aplikasi tata bahasa pada ujaran nyata dalam kajian penampilan bahasa. [?] menggambarkan karakter tata bahasa mengenai hal ini sebagai berikut: frasa yang mengandung banyak konstituen cenderung memiliki percabangan searah, sedangkan arah dependensi frasa yang hanya mengandung satu konstituen ditemukan tidak konsisten. Kajian [?] yang dikutip dalam [?] memperlihatkan dengan bukti empiris adanya gabungan percabangan searah dan percabangan beda arah (*mixed branching*) untuk menghasilkan keseimbangan antara arah dependensi positif (diawali induk) dan dependensi negatif (diakhiri induk) dalam mendapatkan nilai panjang dependensi terkecil (4).

II. METHODS

The written data, obtained from a partnership with an Indonesian technology company called Dattabot, consists of 9311 sentences of various news articles published from

¹Seperti yang dijelaskan sebelumnya, induk (*head*) di sini sedikit berbeda dengan induk (*heads*) pada teori struktur frasa. Sehingga, bentuk relasi dependensi diawali atau diakhiri induk yang dibahas pada bagian ini terbatas hanya pada penerapannya berdasarkan teori dependensi dan bukan struktur frasa.

[width=0.8] pics/balancedbranching.png

Figure 4. Keseimbangan direksionalitas induk

2008 to 2018. The spoken data, obtained from partnerships with various journalists across the country, consists of 10.219 sentences of live reports, interviews, and other spontaneous utterances (all in the journalistic domain) recorded from 2010 to 2018. The main challenge in including spoken data is because computational parsers are mostly trained to parse written data. Therefore, sentences with specific features of spoken language, such as hesitations, fillers, etc., are not included in the corpus. These two sets of corpora are preprocessed, parsed, and cleaned using UDPipe udpipe2017 and manual verifications. UDPipe udpipe2017 is a trainable pipeline for tokenization, tagging, lemmatization and dependency parsing of CoNLL-U files based on dependency treebanks provided by Universal Dependencies 2.0 nivre2017universal that are contextually adjusted for many languages, including Indonesian.

Data jurnalistik ragam tulis yang diterima dari hasil kerja sama dengan perusahaan media monitoring Dattabot berbentuk teks elektronik hasil pemindaian media cetak dengan pendekatan komputasional *Optical Character Recognition* (OCR). Sedangkan, data jurnalistik ragam lisan yang diterima berupa rekaman audio yang selanjutnya diolah terlebih dahulu dalam salah satu tahap persiapan data menggunakan metode komputasional dengan bantuan Google Speech API dan diverifikasi kembali secara manual. Tahap berikutnya memanfaatkan UDPipe udpipe2017 yang merupakan sebuah kerangka kerja yang dapat dilatih ulang untuk pengkategorian konstituen kalimat (*token*), penguraian lemma, dan penguraian kalimat. Kerangka kerja UDPipe digunakan sebagai langkah penguraian kalimat komputasional dan dijalankan sebagai paket CRAN pihak ketiga UDPipe versi 0.5 udpipe2017manual pada bahasa pemrograman R (versi 3.3.3) r2017project. Bahasa pemrograman R ini dijalankan dengan RStudio rstudio2017, sebuah perangkat lunak sumber terbuka.

Umumnya, kerangka kerja atau pendekatan penguraian kalimat dalam ranah linguistik komputasional memanfaatkan teori konstituen sebagai dasar teori dan pengembangannya sehingga menghasilkan uraian kalimat berdasarkan struktur frasanya. Kerangka kerja penguraian kalimat UDPipe menggunakan dasar bank pohon struktur dependensi dari koleksi Universal Dependencies 2.0 nivre2017universal dengan memadukan lemma dan fitur morfologis lainnya dari kerangka kerja MorphInd yang kontekstual terhadap data bahasa Indonesia larasati2011indonesian. Kerangka kerja penguraian kalimat berdasarkan dependensi saat ini masih terus dikembangkan secara intensif sehingga masih memiliki kekurangan dalam aspek menangani keambiguitasan makna konstituen, bentuk bahasa informal, dan masalah tuturan nyata lainnya. Kekurangan ini menuntut tahapan verifikasi manual untuk memastikan kualitas data. Meskipun begitu, hasil uji model terhadap bahasa Indonesia memiliki akurasi 74,3%-80,6% udpipe2017. Setelah melalui proses penguraian kalimat, seluruh tahapan-tahapan berikutnya yang mencakup anotasi nilai DL dan MDD, seleksi dan

klasifikasi data, analisis teks, pemaparan pola DL dan MDD, serta uji signifikansi memanfaatkan bahasa pemrograman R (versi 3.3.3) r2017project dan Python (versi 3.6.3) (Python Core Team, 2017).

Direksionalitas induk berkaitan dengan hierarki relasi antara induk dan konstituen terikatnya. Seperti yang dijelaskan pada bab sebelumnya dan divisualisasikan pada 1, relasi antara dua konstituen dihubungkan dengan sebuah tautan yang memiliki panah. Arah panah tersebut merepresentasikan hierarki relasi antara keduanya. Arah panah ke kanan menandakan tautan yang bernilai positif yang mewakili bentuk relasi diawali induk. Sebaliknya, arah panah ke kiri menandakan tautan yang bernilai negatif yang mewakili bentuk relasi diakhiri induk. Hal ini dapat dilihat juga pada visualisasi dari dua kalimat contoh yang diambil dari data ragam tulis (??) dan ragam lisan (??). Kedua visualisasi ini mengandung tautan-tautan dengan anotasi positif (+) dan anotasi negatif (−) (yang juga dibedakan oleh warna biru dan merah) untuk memberikan gambaran terhadap perbedaan direksionalitas induk dalam sebuah kalimat. Sebagai contoh, ?? memiliki 4 tautan dependensi positif dan 1 tautan dependensi negatif, sehingga pada kalimat ini relasi di mana induk direalisasikan sebelum konstituen terikat mendominasi kalimat tersebut. Pada level dependensi utama atau simpai pusat, *menggelar* memiliki 2 tautan dependensi positif terhadap *program* dan *Cassina* serta 0 tautan dependensi negatif.

Simpai pusat merepresentasikan relasi-relasi utama dalam sebuah kalimat berdasarkan dependensi. Pada kalimat yang lebih kompleks, simpai pusat juga memberikan gambaran terhadap relasi-relasi utama, namun terkadang banyak relasi pada cabang yang tidak terwakili. Oleh karena itu, pemaparan kedua analisis direksionalitas induk pada relasi-relasi secara keseluruhan dan pada simpai pusat penting untuk disampaikan. Pada tahap anotasi DL, setiap relasi antara dua konstituen dalam sebuah kalimat sudah diberikan anotasi positif maupun negatif. Oleh sebab itu, pada tahap ini hanya perlu dilakukan agregasi data terhadap relasi-relasi bernilai positif dan relasi-relasi bernilai negatif. Tahap analisis ini didasarkan pada pembahasan mengenai percabangan searah dan percabangan beda arah yang dibahas oleh [?] dan [?] untuk melihat adanya kecenderungan dalam menerapkan tipe percabangan keduanya. Uji hipotesis untuk bagian analisis ini dilakukan dengan pendekatan penghitungan statistika deskriptif.

-
-
- To analyze the basic comparison between written (formal) and spoken (spontaneous) language, the corpora are first classified into two sets: written data and spoken data.
 - To investigate the possible influence sentence length may have on dependency distances (and syntactic construction). This study considers approaches from similar studies oya2011syntactic, jiang2015effects and classifies both sets of corpora into five categories:
 - Sentences with 5 constituents or less.
 - Sentences with 6 to 10 constituents.

.32 [width=1] pics/ts6841.jpg

Figure 5. *Tumuma*

.46 [width=1]pics/l57375.jpg

Figure 6. *Lumuma*

Figure 7. Contoh perbedaan relasi diawali dan diakhiri induk pada hubungan dua konstituen secara umum

.7 [width=1] pics/tulis_DLposneg.png

Figure 8. Grafik panjang dependensi positif dan negatif data ragam tulis

.7 [width=1]pics/lisan_DLposneg.png

Figure 9. Grafik panjang dependensi positif dan negatif data ragam lisan

Figure 10. Grafik jumlah konstituen (sumbu *x*) terhadap panjang dependensi kalimat (sumbu *y*) untuk relasi diawali induk atau dependensi positif (biru) dan relasi diakhiri induk atau dependensi negatif (merah) korpus data ragam tulis (atas) dan lisan (bawah). Satu kalimat memiliki satu titik yang mewakili panjang dependensi positif dan satu titik yang mewakili dependensi negatif. Garis regresi yang terbentuk merupakan gambaran rata-rata yang didapatkan dari seluruh titik setiap aspek dependensi.

- Sentences with 11 to 20 constituents.
- Sentences with 21 to 30 constituents.
- Sentences with more than 30 constituents.

III. RESULTS

Pada hubungan dua konstituen secara umum, setiap tautan dependensi ikut diperhitungkan dari segi frekuensi maupun jaraknya. Sebagai contoh, 5 dan 6 merupakan dua kalimat yang ditemukan pada korpus data ragam tulis dan lisan. Dalam 5, seluruh tautan dependensinya adalah positif, dalam arti induk menempati posisi sebelum konstituen terikatnya (diawali induk). Kalimat ini memiliki frekuensi dependensi positif sebanyak 5 dengan jarak masing-masing adalah 1, 2, 3, 1, dan 1 serta frekuensi dependensi negatif sebanyak 0. Sebaliknya, 6 hanya memiliki tautan dependensi negatif, dalam arti induk menempati posisi setelah konstituen terikatnya (diakhiri induk). Kalimat ini memiliki frekuensi dependensi positif sebanyak 0 serta frekuensi dependensi negatif sebanyak 5 dengan jarak masing-masing adalah 5, 4, 3, 2, dan 1.

Pada uji hipotesis ini, kecenderungan dilihat dari segi frekuensi dan jarak dependensi. Seperti contoh yang dijabarkan di atas, frekuensi tidak selalu berkorelasi secara linear terhadap jarak sehingga kecenderungan pada satu aspek dapat berbeda dengan aspek lainnya. Meskipun kalimat yang tidak memiliki salah satu bentuk relasi diawali maupun diakhiri induk seperti di atas tidak banyak ditemukan, banyak kalimat dalam kedua korpus data yang didominasi oleh salah satu bentuk sehingga menarik untuk dikaji kecenderungannya.

Berdasarkan III dan III, kedua korpus data menunjukkan garis regresi tautan dependensi positif yang berada jauh di atas garis regresi tautan dependensi negatif. Pada kalimat dengan jumlah konstituen yang semakin banyak, jarak antara keduanya semakin jauh. Seperti yang

Table I
FREKUENSI SELISIH PANJANG DEPENDENSI POSITIF DAN NEGATIF
PADA HUBUNGAN DUA KONSTITUEN SECARA UMUM

Length	Written (+ >-)	Written (+ <-)	Spoken (+ >-)	Spoken (+ <-)
<= 5	164	221	1188	1354
6 - 10	992	646	1469	1349
11 - 20	3145	952	1858	1014
21 - 30	1854	200	644	166
31 - 40	558	36	204	37
>40	201	5	89	7

dijelaskan sebelumnya, panjang dependensi cukup sensitif terhadap panjang kalimat dan keduanya memiliki hubungan linear. Dalam sebuah kalimat, perbandingan jumlah bentuk relasi diawali dan diakhiri induk saling berkaitan. Hal ini berarti ada kecenderungan preferensi terhadap bentuk relasi diawali induk dan menandakan kesesuaian data observasi dengan tata bahasa yang berlaku. Namun, pada grafik arah depensi ragam lisan (III), jarak garis regresi tautan dependensi positif dan negatif terlihat lebih dekat dibandingkan ragam tulis. Grafik ini juga memperlihatkan adanya kurva parabola terbuka ke bawah sehingga muncul asumsi adanya nilai tertinggi atau ambang batas (*threshold*) terhadap panjang dependensi negatif pada korpus data ragam lisan. Berdasarkan korpus data yang terkumpul, dugaan ambang batas tautan dependensi negatif ini menimbulkan asumsi bahwa mulai panjang kalimat tertentu, panjang dependensi negatif dalam sebuah kalimat mungkin tidak akan melebihi nilai tertentu juga.

I berisi frekuensi kalimat dengan melihat apakah kalimat tersebut memiliki panjang dependensi positif yang lebih jauh atau sebaliknya. Hal ini berarti kalimat yang memiliki panjang dependensi positif dan negatif sama jauh tidak diikutsertakan dalam analisis. Jumlah kalimat dalam I dihitung berdasarkan selisih panjang dependensinya. Berdasarkan tabel ini, kalimat yang panjang dependensi positifnya lebih besar dari negatif ditemukan lebih banyak mulai dari panjang kalimat 6 konstituen. Hal ini memperlihatkan adanya kecenderungan terhadap bentuk relasi diawali induk. Selain itu, rasio perbandingan pada ragam tulis jauh lebih besar dibandingkan ragam lisan. Pada kategori kalimat terpanjang, jumlah kalimat yang panjang dependensi positifnya lebih besar dapat mencapai 40 kali lebih banyak. Temuan ini menunjukkan bahwa penutur cenderung lebih banyak menempatkan induk pada posisi sebelum konstituen terikatnya. Seiring bertambahnya jumlah konstituen, rasio frekuensi tautan dependensi positif dan negatif semakin besar pada kedua ragam. Namun, perbedaan rasio dari setiap kategori jumlah konstituen pada ragam lisan tidak sejauh ragam tulis.

II merupakan tabel yang berisi frekuensi tautan dependensi positif dan negatif dalam kedua korpus data. Frekuensi ini hanya menggambarkan jumlah kemunculan tautan dan tidak memiliki jarak dependensi. Pada tabel ini, tautan-tautan dependensi tidak dikelompokkan per kalimat sehingga dapat memberikan gambaran jumlah tautan yang ada dalam keseluruhan korpus. Data ragam tulis menunjukkan bahwa mulai dari panjang kalimat 6 konstituen,

Table II
FREKUENSI TAUTAN DEPENDENSI POSITIF DAN NEGATIF PADA
HUBUNGAN ANTARA DUA KONSTITUEN SECARA UMUM

Length	Written (+)	Written (-)	Spoken (+)	Spoken (-)
<= 5	677	700	2841	2764
6 - 10	7023	5715	7649	6918
11 - 20	35597	24426	15554	12898
21 - 30	31036	18163	7564	5939
31 - 40	12904	6968	3232	2414
>40	6583	3240	1928	1426

Table III
JARAK DEPENDENSI SELURUH TAUTAN ANTARA DUA KONSTITUEN

Length	+				-				
	min	med	max	mean	min	med	max	mean	
<= 5	1	1	4	1,409	1	1	4	1,606	6*Written
6 - 10	1	1	9	1,866	1	1	9	1,948	
11 - 20	1	1	19	2,468	1	1	19	2,147	
21 - 30	1	1	29	2,971	1	1	29	2,241	
31 - 40	1	1	39	3,597	1	1	37	2,299	
>40	1	1	68	4,22	1	1	55	2,282	6*Spoken
<= 5	1	1	4	1,466	1	1	4	1,534	
6 - 10	1	1	9	1,933	1	1	9	2,071	
11 - 20	1	2	19	2,535	1	1	19	2,321	
21 - 30	1	2	28	3,263	1	1	27	2,5	
31 - 40	1	2	37	3,718	1	1	32	2,657	
>40	1	2	66	4,067	1	1	46	2,438	

frekuensi tautan dependensi negatif lebih sedikit dibandingkan tautan dependensi positif. Meskipun begitu, tautan dependensi positif lebih banyak pada semua klasifikasi dalam ragam lisan.

Pada kategori kalimat terpanjang dalam ragam tulis, frekuensi tautan dependensi positif dapat mencapai dua kali tautan dependensi negatif. Sementara pada ragam lisan, perbandingan terbesar hanya mencapai 1,35 banding 1 sehingga juga menunjukkan bahwa kecenderungan bentuk relasi diawali induk lebih terlihat pada data ragam tulis. Rasio dengan jumlah tautan dependensi pada tabel ini memperlihatkan perbandingan yang tidak seekstrim rasio dengan jumlah kalimat pada tabel sebelumnya. Temuan ini memperkuat temuan sebelumnya bahwa penutur bahasa Indonesia cenderung menempatkan posisi induk sebelum konstituen terikatnya. Temuan ini juga menunjukkan bahwa panjang dependensi dari tautan dependensi positif dapat memiliki nilai yang jauh lebih besar terlepas dari perbandingan jumlahnya terhadap tautan dependensi negatif dalam satu kalimat.

III mencakup jarak dependensi dari seluruh tautan yang ada dalam kedua korpus data dan tidak dikelompokkan berdasarkan kalimatnya. Tabel ini mendukung temuan sebelumnya dan memperlihatkan bahwa rata-rata jarak dependensi positif lebih jauh dibandingkan dependensi negatif, terutama pada kalimat yang semakin panjang. Perbandingan ini mulai terlihat pada kategori panjang kalimat 11 konstituen dan meningkat seiring bertambahnya jumlah konstituen. Pada kategori kalimat terpanjang, perbandingan jarak antara dependensi positif dan negatif juga semakin jauh. Pada III, terlihat bahwa hampir semua nilai tengah jarak dependensi sama atau sangat mendekati nilai minimum terutama pada ragam tulis. Hal ini menunjukkan

Table IV
RATA-RATA JARAK DEPENDENSI POSITIF DAN NEGATIF SELURUH
TAUTAN ANTARA DUA KONSTITUEN

Length	+				-			
	min	med	max	mean	min	med	max	mean
<= 5	1	1	3,5	1,372	1	1,5	4	1,545
6 - 10	1	1,667	5,2	1,839	1	1,667	9	1,895
11 - 20	1	2,25	7,25	2,432	1	1,875	10	2,117
21 - 30	1	3,765	10,278	2,97	1	2	12,75	2,212
31 - 40	1,458	3,368	9,933	3,595	1	2	8	2,272
>40	1,929	3,756	10,567	4,154	1	2	7,111	2,258
<= 5	1	1	4	1,378	1	1,33	4	1,456
6 - 10	1	1,714	5,6	1,867	1	1,8	9	1,903
11 - 20	1	2,286	9,625	2,479	1	2	6,818	2,255
21 - 30	1,4	3	8,412	3,252	1	2,188	8,636	2,447
31 - 40	1,556	3,512	7,25	3,719	1	2,27	8,803	2,612
>40	2,174	3,865	6,895	3,957	1,059	2,294	4,2	2,363

bahwa ada kecenderungan banyaknya dua konstituen yang saling berdampingan sehingga menghasilkan nilai minimum atau 1. Pada kategori kalimat terpanjang dalam kedua ragam (>40 konstituen), rata-rata jarak dependensi negatif lebih pendek dibandingkan kategori sebelumnya (31-40 konstituen). Meskipun demikian, perbedaan yang signifikan ($P < 0.05$) hanya ditemukan pada ragam lisan. Hal ini berarti pada ragam lisan, mulai panjang kalimat tertentu (yang berdasarkan data ini adalah sekitar 40 konstituen) jarak dependensi negatif tidak akan melebihi nilai tertentu.

Nilai-nilai pada IV didapatkan dengan menghitung rata-rata jarak dependensi positif dan negatif pada setiap kalimat. Distribusi nilai minimum, tengah dan maksimum pada ragam lisan lebih merata dibandingkan ragam tulis yang nilai tengahnya terlihat lebih mendekati nilai minimum terutama pada kalimat yang semakin panjang (mulai terlihat pada panjang kalimat 11 konstituen). Mulai panjang kalimat tersebut, rata-rata jarak dependensi pada ragam tulis juga terlihat lebih pendek. Hal ini mendukung temuan sebelumnya bahwa dibandingkan ragam lisan, bahasa Indonesia ragam tulis lebih memperlihatkan upaya untuk meningkatkan efisiensi kalimat dengan mendekatkan konstituen-konstituen yang memiliki tautan dependensi sehingga menghasilkan panjang dan jarak dependensi yang lebih pendek terutama pada kalimat yang lebih panjang. Seperti III, tabel ini juga memperlihatkan bahwa tautan dependensi negatif memiliki rata-rata jarak yang lebih pendek. Rasio perbandingan dependensi positif dan negatif pada ragam lisan juga tidak sebesar data ragam tulis.

A. Bentuk relasi diakhiri induk cenderung berjarak lebih pendek pada simpai pusat

Sesuai dengan batasan penelitian ini, bagian analisis ini difokuskan hanya pada simpai pusat dengan kelas kata akar berupa kata kerja atau verba. 11 dan 12 merupakan dua kalimat yang ditemukan pada korpus data ragam lisan. 13 menampilkan contoh perbedaan penempatan klausa terikat *kalau boleh tahu*. Klausa terikat *kalau boleh tahu* pada kalimat *Lakara* (11) dan *Lakarb* (12) berada pada posisi yang berbeda namun tidak mengubah makna kalimat. Contoh serupa yang menggambarkan perbedaan

.4 [width=1] pics/lsl436.jpg

Figure 11. *Lakara*

.58 [width=1]pics/lsl460.jpg

Figure 12. *Lakarb*

Figure 13. Contoh perbedaan penempatan klausa pada data ragam lisan

.7 [width=1] pics/tulisroot_DLposneg.png

Figure 14. Grafik panjang dependensi positif dan negatif simpai pusat data ragam tulis

.7 [width=1]pics/lisanroot_DLposneg.png

Figure 15. Grafik panjang dependensi positif dan negatif simpai pusat data ragam lisan

Figure 16. Grafik jumlah konstituen (sumbu x) terhadap panjang dependensi simpai pusat akar verbal (sumbu y) untuk relasi diawali induk atau dependensi positif (biru) dan relasi diakhiri induk atau dependensi negatif (merah) korpus data ragam tulis (atas) dan lisan (bawah). Satu simpai pusat memiliki satu titik yang mewakili panjang dependensi positif dan satu titik yang mewakili dependensi negatif. Garis regresi yang terbentuk merupakan gambaran rata-rata yang didapatkan dari seluruh titik setiap aspek dependensi.

letak klausa terikat atau perbedaan urutan kata juga sempat dijabarkan oleh [hlm. 209]sneddon2010indonesian dalam buku *Indonesian Reference Grammar*.

Pada kalimat *Lakarb*, konstituen-konstituen yang membentuk klausa *kalau boleh tahu* secara kolektif memiliki relasi dependensi utama negatif terhadap akar *ordernya*. Sesuai dengan bentuk relasi induk dan konstituen terikat berdasarkan teori dependensi yang dijabarkan oleh [?], klausa terikat pada posisi awal kalimat ini harus disimpan dalam memori kerja terlebih dahulu hingga akar direalisasikan. Karena akar mengandung informasi utama, perbedaan kedua kalimat ini juga dapat memberikan sedikit gambaran bagaimana perbedaan proses memori kerja terkait penyimpanan informasi yang merupakan argumen inti kalimat.

Table V
FREKUENSI SELISIH PANJANG DEPENDENSI POSITIF DAN NEGATIF
PADA SIMPAI PUSAT AKAR VERBAL

Length	Written (+ >-)	Written (+ <-)	Spoken (+ >-)	Spoken (+ <-)
<= 5	57	180	393	1008
6 - 10	475	796	655	1356
11 - 20	1638	1804	964	1336
21 - 30	1028	778	384	291
31 - 40	325	177	121	85
>40	137	37	61	30

Table VI
FREKUENSI TAUTAN DEPENDENSI POSITIF DAN NEGATIF PADA
SIMPAI PUSAT AKAR VERBAL

Length	Written (+)	Written (-)	Spoken (+)	Spoken (-)
<= 5	201	412	1204	2151
6 - 10	1917	2532	2998	4433
11 - 20	6914	7144	4753	5570
21 - 30	4455	3691	1869	1754
31 - 40	1528	1031	540	461
>40	625	321	315	258

III-A dan III-A merupakan grafik panjang dependensi positif dan negatif yang didapat pada simpai pusat kalimat dengan akar berupa verba. Kalimat dengan akar verbal ditemukan sebanyak 84,61% atau sebanyak 7874 kalimat pada data ragam tulis dan 70,91% atau sebanyak 7239 kalimat pada data ragam lisan. Mayoritas frekuensi kemunculan akar verbal ini menjadi dasar untuk melihat bagaimana penutur menyusun informasi utama dengan meninjau tautan-tautan dependensi utama (simpai pusat). Oleh sebab itu, semua klausa yang mungkin terikat pada simpai cabang dianggap mengikuti induknya seperti pada contoh uraian relasi klausa *kalau boleh tahu* dan akarnya di atas (13). 16 menunjukkan adanya konsistensi pada tingkat tautan dependensi utama dan keseluruhan kalimat. Grafik ini memperlihatkan bahwa penutur juga cenderung memilih untuk menekan penggunaan bentuk relasi diakhiri induk terutama pada kalimat panjang. Berbeda dengan 10 yang pada klasifikasi kalimat menengah sudah cukup terlihat perbedaan antara panjang dependensi positif dan negatif, grafik ini menunjukkan panjang dependensi positif dan negatif saling tumpang tindih sehingga perlu ditinjau lebih dalam. Meskipun begitu, serupa dengan 10, garis regresi panjang dependensi negatif mengindikasikan adanya ambang batas pada nilai tertentu terutama pada data ragam lisan.

V berisi jumlah-jumlah kalimat dengan melihat apakah kalimat tersebut memiliki panjang dependensi positif yang lebih jauh atau sebaliknya pada simpai pusat yang memiliki akar verbal. Berbeda dengan I, kalimat yang panjang dependensi positifnya lebih jauh ditemukan lebih banyak pada kalimat-kalimat panjang (mulai 21 konstituen). Pada ragam tulis, frekuensi keseluruhan hampir seimbang sedangkan pada ragam lisan cukup berbeda jauh (total frekuensi di mana panjang dependensi negatif lebih jauh dari panjang dependensi positif sebanyak dua kali lipat dibandingkan sebaliknya).

VI memperlihatkan bahwa pada simpai pusat akar

Table VII
JARAK DEPENDENSI SELURUH TAUTAN ANTARKONSTITUEN PADA
SIMPAI PUSAT AKAR VERBAL

Length	+				-				
	min	med	max	mean	min	med	max	mean	
<= 5	1	1	3	1,428	1	1	4	1,748	6*Tulis
6 - 10	1	2	9	2,336	1	2	9	2,571	
11 - 20	1	3	18	4,083	1	2	19	3,61	
21 - 30	1	4	27	6,265	1	3	29	4,793	
31 - 40	1	6	38	9,046	1	3	37	5,925	
>40	1	9	68	12,88	1	3	44	6,791	
<= 5	1	1	4	1,482	1	1	4	1,69	6*Lisan
6 - 10	1	2	9	2,28	1	2	9	2,618	
11 - 20	1	3	19	3,826	1	2	18	3,658	
21 - 30	1	4	26	6,413	1	3	27	4,789	
31 - 40	1	6	35	8,638	1	4	32	6,505	
>40	1	7	66	10,27	1	3	46	5,852	

verbal, kecenderungan menempatkan induk pada posisi sebelum konstituen terikatnya juga baru terlihat pada panjang kalimat mulai 21 konstituen. Pada kalimat-kalimat yang lebih pendek dari itu, frekuensi tautan dependensi negatif yang merepresentasikan bentuk relasi diakhiri induk justru lebih banyak. Pada data ragam tulis, total frekuensi dependensi negatif juga lebih banyak dibandingkan dependensi positif namun rasionya tidak seekstrim pada V.

VII mencakup jarak dependensi dari seluruh tautan pada simpai pusat akar verbal dan tidak dikelompokkan berdasarkan kalimatnya. Tabel ini memperlihatkan bahwa rata-rata jarak dependensi positif lebih jauh dibandingkan dependensi negatif, terutama pada kalimat yang semakin panjang. Seperti pada III, perbandingan ini mulai terlihat pada kategori panjang kalimat 11 konstituen dan meningkat seiring bertambahnya jumlah konstituen. Pada kategori kalimat terpanjang, perbandingan jarak antara dependensi positif dan negatif juga semakin jauh. Keserupaan temuan pada simpai pusat akar verbal dan pada hubungan dua konstituen secara umum (III) menunjukkan bahwa meskipun frekuensi dependensi negatif lebih banyak (terutama pada ragam lisan), tautan dependensi negatif ditemukan berjarak lebih pendek. Nilai tengah untuk semua ragam dan kategori panjang kalimat juga lebih mendekati ke nilai minimum, terutama pada ragam tulis. Pada kategori kalimat terpanjang ragam lisan (>40 konstituen), rata-rata jarak dependensi negatif lebih pendek dibandingkan kategori sebelumnya (31-40 konstituen) dan perbedaan ini signifikan ($P < 0.05$). Serupa dengan III, mulai panjang kalimat tertentu (yang berdasarkan data ini adalah sekitar 40 konstituen), jarak dependensi negatif pada simpai pusat ragam lisan juga tidak akan melebihi nilai tertentu.

Hasil pada VIII didapatkan dengan menghitung rata-rata jarak dependensi positif dan negatif pada simpai pusat akar verbal setiap kalimat. Seperti IV, distribusi nilai minimum, tengah dan maksimum pada ragam lisan juga lebih merata dibandingkan ragam tulis yang nilai tengahnya terlihat lebih mendekati nilai minimum terutama pada kalimat yang semakin panjang (mulai terlihat pada panjang kalimat 11 konstituen). Namun, perbedaannya terletak pada rata-rata jarak dependensi. Pada simpai pusat

Table VIII
RATA-RATA JARAK DEPENDENSI POSITIF DAN NEGATIF PADA SIMPAI
PUSAT AKAR VERBAL

Length	+				-				
	min	med	max	mean	min	med	max	mean	
<= 5	1	1	3	1,363	1	1,5	4	1,632	6*Tulis
6 - 10	1	2	7	2,091	1	2	9	2,385	
11 - 20	1	3	16	3,545	1	2,75	19	3,36	
21 - 30	1	5	19,5	5,514	1	3,333	28	4,442	
31 - 40	1	7,5	26	8,044	1	4	36	5,448	
>40	1	11,1	34,286	11,84	1	4,2	42	6,264	
<= 5	1	1	3	1,327	1	1,5	4	1,531	6*Lisan
6 - 10	1	1,667	5,6	1,801	1	2	9	2,603	
11 - 20	1	2,286	7,167	2,479	1	2,125	6,818	2,44	
21 - 30	1,4	3,118	8,412	3,375	1	2,182	8,636	2,688	
31 - 40	1,947	3,643	7,25	3,879	1	2,353	7,5	2,47	
>40	2,147	4,051	8,697	4,156	1,059	2,357	5,053		

ini, rata-rata jarak dependensi ragam tulis jauh lebih besar dibandingkan ragam lisan. Seperti VII, tabel ini juga memperlihatkan bahwa tautan dependensi negatif memiliki rata-rata jarak yang lebih pendek untuk kedua ragam.

B. Percabangan searah dan keseimbangan antara tautan dependensi positif dan negatif

Pada tinjauan pustaka, pergerakan panah yang merepresentasikan direksionalitas induk dibahas menjadi dua tipe, yaitu percabangan searah (*same-branching*) dan percabangan beda arah (*mixed-branching*) (hawkins1994performance, frazier1985syntactic). Namun, dengan jumlah konstituen yang semakin banyak dan kalimat yang semakin kompleks, direksionalitas induk tidak bisa digolongkan secara biner (hanya percabangan searah saja atau hanya percabangan beda arah saja) (dryer1992greenbergian, temperley2008dependency). [?] menjelaskan mengenai konsep keseimbangan antara tautan dependensi positif dan negatif dalam mendapatkan panjang dependensi terpendek yang juga dianalisis oleh dryer1992greenbergian. Berdasarkan data penelitian ini, temuan yang didapatkan mendukung asumsi yang dikemukakan [?] bahwa arah dependensi tidak bisa digolongkan secara biner pada tataran kalimat, terutama pada kalimat yang semakin panjang. Namun, berdasarkan analisis dengan pemetaan kepadatan selisih antara panjang dependensi positif dan negatif, belum ditemukan kecenderungan terhadap bentuk keseimbangan tersebut.

Pada ragam tulis, pemetaan kepadatan sangat menyebar sehingga tidak membentuk pola tertentu, sedangkan pada ragam lisan, pemetaan kepadatan terlalu memusat pada kalimat dengan jumlah konstituen sedikit sehingga juga tidak membentuk pola tertentu. Untuk mencari kecenderungan pada data seperti ini, perlu dilakukan pencarian fungsi matematis untuk menguji tingkat kesalahan (*error rate*) yang lebih kecil antara kedua kemungkinan (percabangan searah atau keseimbangan dependensi positif dan negatif). Namun, pencarian fungsi matematis tersebut berada di luar batasan penelitian ini dan di luar ranah linguistik sehingga perlu dilakukan analisis lebih lanjut dari segi pembentukan model statistika. Asumsi sementara yang dapat diambil bertitik tolak dari grafik perbandingan panjang dependensi positif dan negatif pada 10 dan 16

yang dapat memberikan gambaran adanya kecenderungan untuk menghindari tautan dependensi negatif, terutama pada kalimat yang semakin panjang.

IV. CONCLUSION

direksionalitas induk dapat menyebabkan nilai dependensi tersebut menjadi positif atau negatif. Anotasi positif menandakan bentuk relasi induk sebelum konstituen terikat (diawali-induk), sedangkan anotasi negatif menandakan bentuk relasi induk setelah konstituen terikat (diakhiri induk). Berdasarkan aturan struktur frasa dalam tata bahasa yang ada (kridalaksana2002struktur, sneddon2010indonesian), bahasa Indonesia tergolong bahasa yang memilih bentuk relasi diawali induk dibandingkan diakhiri induk. Namun, belum ada penelitian dengan skala cukup besar yang dapat memberikan informasi ini dengan memanfaatkan data ujaran nyata karena penerapannya dalam ujaran nyata dapat bersifat tidak gramatikal namun tetap diterima. Bentuk relasi diawali induk tidak selalu menjadi preferensi pada semua bahasa karena beberapa bahasa memiliki aturan tata bahasa yang menuntut induk muncul setelah konstituen terikat atau diakhiri induk. Meskipun belum ada konvensi dan bukti empiris yang menunjukkan bahwa bentuk diawali induk lebih memudahkan proses memori kerja, [?] menunjukkan bukti empiris bahwa bahasa-bahasa yang cenderung menggunakan bentuk relasi ini memiliki tingkat pengurangan panjang dependensi lebih tinggi dibandingkan bahasa-bahasa yang cenderung menggunakan bentuk relasi diakhiri induk.

Didukung oleh hasil penelitian ini, bahasa Indonesia menunjukkan adanya kecenderungan dalam bahasa Indonesia bahwa induk menempati posisi sebelum konstituen terikatnya pada hubungan antara dua konstituen yang memiliki tautan langsung secara umum. Temuan ini ditandai oleh konsistensi frekuensi dan jarak dari panjang dependensi positif yang semakin besar dibandingkan panjang dependensi negatif terutama pada kalimat yang memiliki jumlah konstituen lebih banyak. Namun, pada argumen utama dengan akar verbal (yang terlihat pada simpai pusat), tidak terlihat ada kecenderungan induk menempati posisi sebelum konstituen terikatnya. Bentuk relasi diakhiri induk lebih banyak ditemukan pada kalimat dengan jumlah konstituen kurang dari 20. Sebaliknya, bentuk relasi diawali induk lebih banyak ditemukan pada kalimat yang panjang. Hal ini dapat dikaitkan dengan salah satu karakter bahasa Indonesia yang memiliki urutan kata bebas (stack2005word, postman2004processing). Hal ini menyebabkan dalam kondisi tertentu, urutan konstituen tersebut dapat berubah dan kalimat masih dapat diterima [pp. 209-268]sneddon2010indonesian.

Posisi induk terhadap konstituen terikat juga berkaitan dengan pola percabangan. Berdasarkan pemetaan dan visualisasi, kecenderungan bahasa Indonesia memanfaatkan strategi percabangan searah atau memanfaatkan strategi

menyeimbangkan dependensi positif dan negatif belum terpapar dengan jelas. Penelitian lebih lanjut yang melibatkan pencarian tingkat kesalahan (*margin of error*) dengan fungsi matematis perlu dilakukan untuk menjawab pertanyaan ini. Meskipun begitu, analisis kualitatif terhadap kalimat-kalimat panjang dalam ragam tulis memperlihatkan banyaknya relasi dependensi antarinduk yang bergerak menuju satu arah tertentu, sehingga menunjukkan adanya indikasi bahwa kalimat-kalimat panjang dalam bahasa Indonesia ragam tulis menerapkan strategi percabangan searah.

=====

ACKNOWLEDGMENT

The authors would like to thank... more thanks here [1]

REFERENCES

- [1] G. K. Zipf, *Human behaviour and the principle of least-effort*. Cambridge, Massachusetts: Addison-Wesley Press, 1949.