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;**

==========
==========
reference figure and tab di dlm text, betulin rumus, fix head-first/head-initial
==========
==========

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 [8]. 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 [1]. Heringer proposed a linear distance that represents dependency between two words or constituents, which has been adopted into modern computational methods of quantifying dependency [2]. 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 [8]. The term *dependency* itself is a derivative of a connection that refers to an asymmetrical relation between a superordinate and its subordinate [**?**], [2], [8]. Many studies and theory developments have all described in detail what a dependency structure between two constituents looks like [8], [**?**], [**?**]. 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 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,

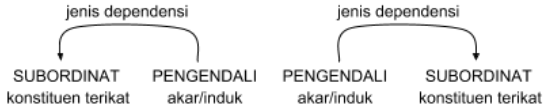

Figure 1. Elements of a dependency relation

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 [**?**], [**?**]. 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 [**?**], [**?**], [4].

Previous substantial studies on measuring dependency used dependency length (DL) as their main approach [3], [4]. 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), which divides DL with the number of dependency relations of an entire sentence [**?**], [**?**]. MDD can be defined as follows:

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

B. Head directionality

There has yet a convention on head directionality in regards to the online processing in working memory, particularly for Indonesian language. However, previous attempts have been done to provide details on the head's position and its relation to grammars. Several studies in the early development of dependency theory found that a language tends to apply a consistent dependency direction, whether it is *head-first*/*head-initial* or *head-last*/*head-final* [**?**], [**?**], [**?**]. Hawkins and Frazier assumed that this consistency, as shown in 2, is the implementation of a strategy to minimize the distance between a head and

Figure 2. Same-branching dependency direction [6]

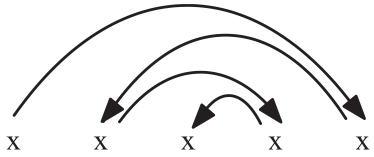

Figure 3. Mixed-branching dependency direction [6]

its dependent [**?**], [**?**].

Temperley argues that a consistent dependency direction or *same-branching* is not an ideal representation of how head directionality works in real utterances [6]. This argument supports Dryer's research who found that the conventional view of same branching only applies on phrases consist of many constituents, while the dependency direction of phrases with single or less words are inconsistent [7]. Gildea and Temperley mentioned this research and provided empirical evidence that exhibit a combination of same branching and mixed branching (3) dependency direction to produce a more balanced head-initial and head-final mixture to yield the shortest dependency distances (4) [3].

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 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 [**?**] and manual verifications. UDPipe 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 that are contextually adjusted for many languages, including Indonesian [**?**], [**?**]. This study uses third party CRAN package UDPipe (version 0.5) on R programming language (version 3.3.3) [**?**], [**?**].

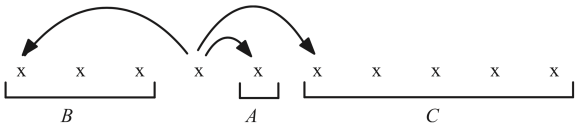

Figure 4. Head-directionality balance [6]

The parsed data are annotated based on the parsing index to measure the dependency distances, length, and mean dependency distances. Ini the parsing index, $root$ has an index of 0. The indexing for the remaining words starts from 1. As mentioned above on Heringer's approach of measuring dependency, this study also uses the number of constituents involved in a dependency relation between two directly-linked constituents as a measuring unit. Therefore, adjacent and directly-linked constituents produce a dependency distance of 1. Data classification, subsetting, and dependency measures are performed on R and Python programming language version 3.3.3 and 3.6.3 respectively [**?**], [**?**]. As visualized in 1, a dependency relation is illustrated through an arc with an arrow. The arrow or direction represents the hierarchy between linked constituents. For Indonesian language, an arrow pointing to the right illustrates a positive $(+)$ dependency relation and denotes a head-initial dependency direction. On the other hand, an arrow pointing to the left illustrates a negative $(-)$ dependency relation and denotes a head-final dependency direction.

A central node in a sentence suggests the main argument or relations of the sentence based on dependency. In a more complex sentence, a branch in a central node can contain multiple constituents. For this reason, the head directionality analysis is performed on two types of relation: (i) on directly-linked constituents in all levels of a dependency tree, including central nodes and adjacent relations, and (ii) on central nodes only, assuming all constituents under each branch are represented by the main arguments. Considering the possible influence sentence length may have on dependency distances, this study classifies both spoken and written data into five categories to acquire more accurate measures [**?**], [**?**], [**?**], [**?**]:

- Sentences with 5 constituents or less.
- Sentences with 6 to 10 constituents.
- Sentences with 11 to 20 constituents.
- Sentences with 21 to 30 constituents.
- Sentences with more than 30 constituents.

III. Results

Indonesian language is considered a free-word order language [**?**]. However, Sneddon describes the many word order rules that put modifiers after their heads based on the functional perspective of phrase structure [**?**]. This study considers this functional view and uses it as a basis to see whether the tendency applies in a dependency structure. This analysis sets apart the head-first (positive) with the head-final (negative) dependency directions to capture a comprehensive characteristic based on the sentence length classifications.

A. Head-initial preference on all levels of a dependency structure

This part of analysis measures dependency distance and number of appearance between head-initial and head-final dependency directions on all levels of a dependency structure. A full consistency of direction such as shown in

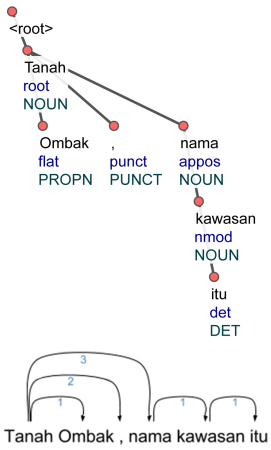

Figure 5. Sentence with all head-initial dependency directions

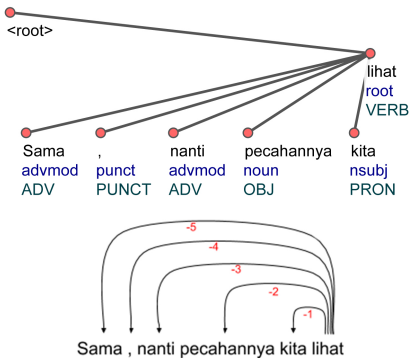

Figure 6. Sentence with all head-final dependency directions

5 and 6 are found only in shorter sentences. 5 illustrates a sentence with all head-initial dependency directions, in contrast to 6. However, the number of appearance between two speech modes is too few to be considered as a preference.

As shown in 7 and 8, both speech modes show preference towards the head-initial dependency direction. The ration difference is increasingly larger in longer sentences. There is even an indication of a threshold to reduce the distance of head-final dependency in the spoken data 8. As seen in table I and II, there are more sentences with longer head-initial dependency length and more head-initial dependencies starting from 6 constituents. The difference between both directions are increasingly larger particularly in longer sentences. However, the margins of difference are significantly larger in the written data.

Based on descriptive statistics using both methods of dependency length (III) and mean dependency distance (IV), preference for head-initial dependencies occur mostly in adjacent relations. This finding support the application of grammar as mentioned by Sneddon who describes that many word order rules put modifiers after their heads based on the functional perspective of phrase structure [**?**]. Both methods show that head-final dependencies are relatively shorter than head-first dependencies in longer sentences. While in shorter sentences, the length

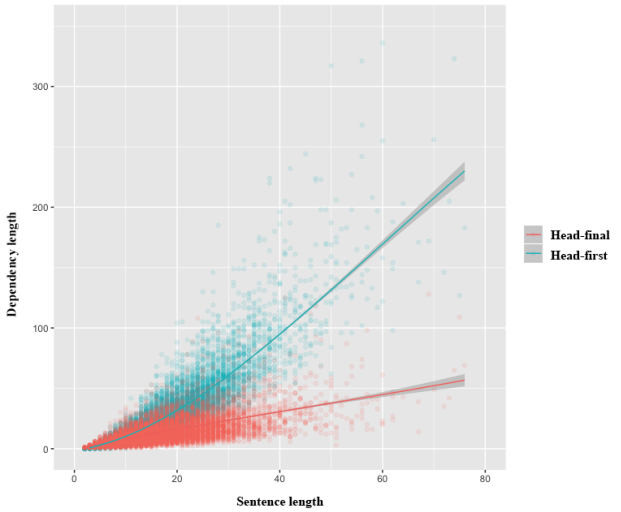

Figure 7. Comparison of head-first and head-final dependencies on all levels based on its length in written data

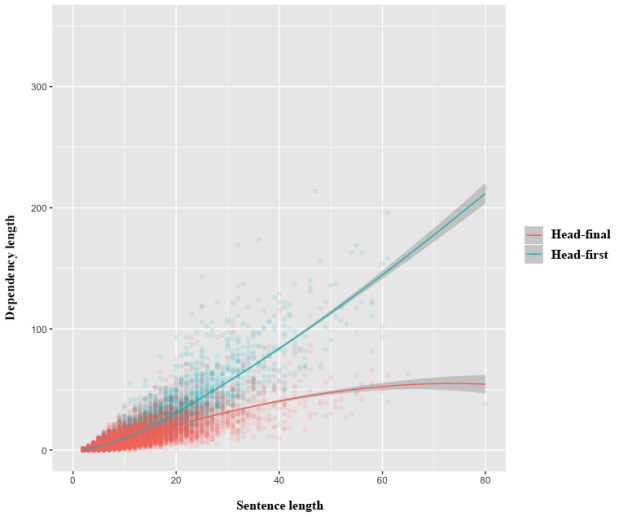

Figure 8. Comparison of head-first and head-final dependencies on all levels based on its length in written data

and distance between both directions are quite similar.

B. Minimization of head-final dependency lengths and distances in longer sentences on central node level

The second part of analysis measures dependency distance and number of occurrence between head-initial and head-final dependencies on central node level. 9 and 10

Table I
NUMBER OF OCCURRENCE WHERE ONE DEPENDENCY DIRECTION IS MORE DOMINANT THAN THE OTHER ON ALL LEVELS OF A DEPENDENCY STRUCTURE

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

Table II
NUMBER OF OCCURRENCE FOR EACH DEPENDENCY DIRECTION ON ALL LEVELS OF A DEPENDENCY STRUCTURE

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	
<= 5	1	1	4	1,466	1	1	4	1,534	6*Spoken
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	

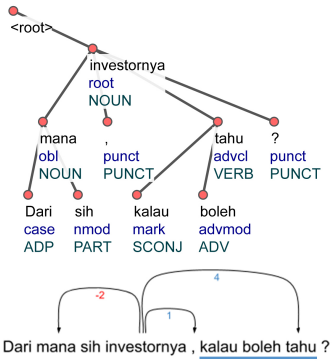

Figure 9. Sentence with a final dependent clause

are two sentences found in the written data that have the same dependent clause *kalau boleh tahu* or "if i may know" in English. Even though each sentence has a different position for this dependent clause, there is no difference in terms of their meanings. The similar example also mentioned by Sneddon in the Indonesian Reference Grammar as one of the key feature of Indonesian as a free-word order language [**?**].

In 9, all of the constituents forming the clause *kalau boleh tahu* collectively creating a head-initial dependency against its head *investornya* or "the investor" in English. On the other hand, the dependent clause in 10 creates a head-final dependency against its head *ordernya* or "the order" in English. In a dependency theory, both constituents forming a dependency must be stored in the working memory until both are realized [8]. In this head-final dependency case, the constituents in this dependent clause (or branch of a dependency structure) must be stored as a whole until the head is realized.

Central node level comprises of constituents that construct the main argument of a sentence. This also relates to how the root is able to bind other constituents, or its valency. To produce a more accurate overview of a main argument in a sentence, this study focuses the analysis on central node level with verbal roots. Sentences with verbal roots are found around 84,61% or 7874 sentences in written data and 70,91% or 7239 sentences in spoken data, which makes it the most common root type in the data.

As opposed to the analysis on all levels of a dependency structure, the preference to either dependency direction is not as clear, as seen in 11 and 12.

As shown in 7 and 8, both speech modes show preference towards the head-initial dependency direction. The difference is increasingly larger in longer sentences. There is even an indication of a threshold to reduce the distance of head-final dependency in the spoken data 8. As seen in table I and II, there are more sentences with longer head-initial dependency length and more head-initial dependencies starting from 6 constituents. The difference between both directions are increasingly larger particularly in longer sentences. However, the margins of difference are significantly larger in the written data.

Based on descriptive statistics using both methods of dependency length (III) and mean dependency distance (IV), preference for head-initial dependencies occur mostly in adjacent relations. This finding support the application of grammar as mentioned by Sneddon who describes that many word order rules put modifiers after their heads based on the functional perspective of phrase structure [**?**]. Both methods show that head-final dependencies are relatively shorter than head-first dependencies in longer sentences. While in shorter sentences, the length and distance between both directions are quite similar.

11 dan 12 merupakan grafik panjang dependensi positif dan negatif yang didapat pada simpai pusat kalimat dengan akar berupa verba. 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

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	2,211
>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	2,003
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	2,565

6*Lisan

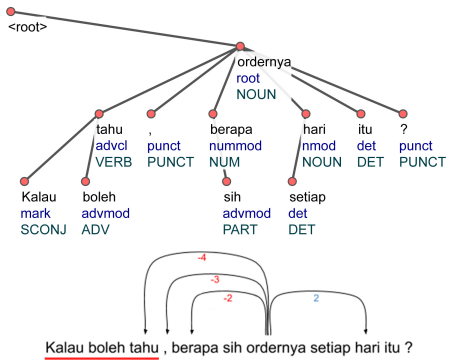

Figure 10. Sentence with an initial dependent clause

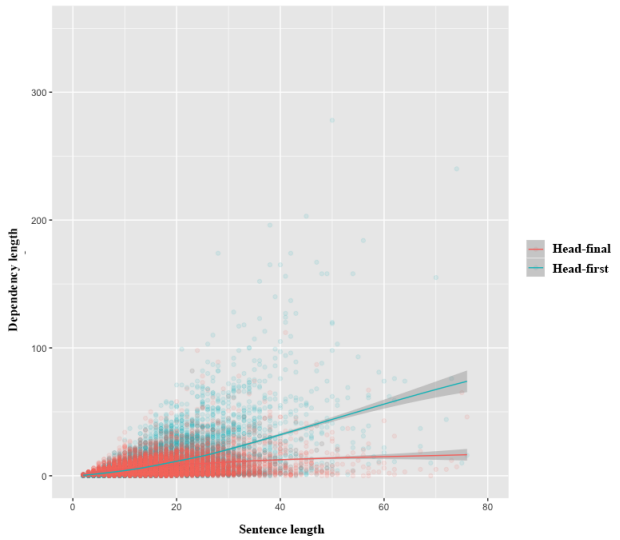

Figure 11. Comparison of head-first and head-final dependencies on central node level in written data

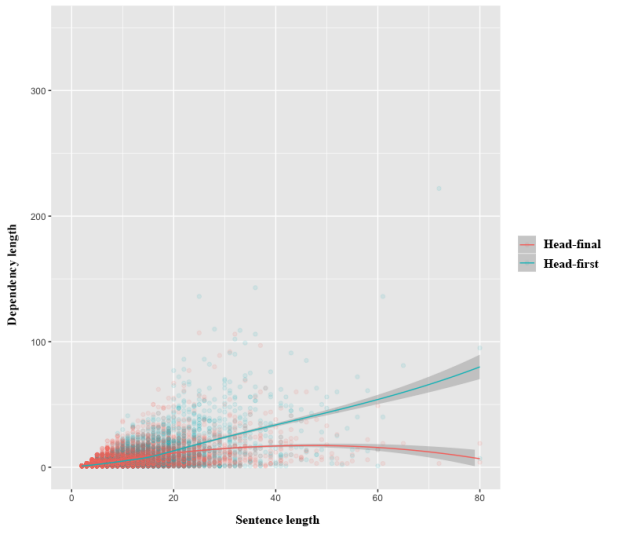

Figure 12. Comparison of head-first and head-final dependencies on central node level in spoken data

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

dianggap mengikuti induknya seperti pada contoh uraian relasi klausa *kalau boleh tahu* dan akarnya di atas (**??**). **??** 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 **??** 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 **??**, 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 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

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 - 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 - 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

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

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

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 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.

C. 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-brancing*) 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). [3] 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 [7] 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 **??** dan **??** 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, [4] 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 [9]

REFERENCES

[1] J. B. Plotkin and M. A. Nowak, "Language evolution and information theory," *Journal of Theoretical Biology*, vol. 205, no. 1, pp. 147–159, 2000.

[2] H. J. Heringer, "Dependency syntax-basic ideas and the classical model," *Syntax-An International Handbook of Contemporary Research*, vol. 1, pp. 298–316, 1993.

[3] D. Gildea and D. Temperley, "Do grammars minimize dependency length?" *Cognitive Science*, vol. 34, no. 2, pp. 286–310, 2010.

[4] R. Futrell, K. Mahowald, and E. Gibson, "Large-scale evidence of dependency length minimization in 37 languages," *Proceedings of the National Academy of Sciences*, vol. 112, no. 33, pp. 10 336–10 341, 2015.

[5] H. Liu, "Dependency direction as a means of word-order typology: A method based on dependency treebanks," *Lingua*, vol. 120, no. 6, pp. 1567–1578, 2010.

[6] D. Temperley, "Dependency-length minimization in natural and artificial language," *Journal of Quantitative Linguistics*, vol. 15, no. 3, pp. 256–282, 2008.

[7] M. S. Dryer, "The Greenbergian word order correlations," *Language*, vol. 68, pp. 81–138, 1992.

[8] L. Tesnière, *Éléments de syntaxe structurale*. Paris, France: Librairie C. Klincksieck, 1959.

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