

Which annotation scheme is more expedient to measure syntactic difficulty and cognitive demand?

Jianwei Yan
Department of Linguistics
Zhejiang University, China
jwyan@zju.edu.cn

Haitao Liu
Department of Linguistics
Zhejiang University, China
lhtzju@gmail.com

Abstract

This paper investigates which annotation scheme of dependency treebank is more congruent for the measurement of syntactic complexity and cognitive constraint of language materials. Two representatives of semantic- and syntactic-oriented annotation schemes, the Universal Dependencies (UD) and the Surface-Syntactic Universal Dependencies (SUD), are under discussion. The results show that, on the one hand, natural languages based on both annotation schemes follow the universal linguistic law of Dependency Distance Minimization (DDM); on the other hand, according to the metric of Mean Dependency Distances (MDDs), the SUD annotation scheme that accords with traditional dependency syntaxes are more expedient to measure syntactic difficulty and cognitive demand.

1 Background and Motivation

Dependency grammar deals with the syntactically related words, i.e. the governor and the dependent, within sentence structure (Hinger, 1993; Hudson, 1995; Liu, 2009). It can be dated back to the seminal work of *Eléments de Syntaxe Structurale* by Tesnière (1959), and developed through different theories, including Word Grammar, Meaning-Text-Theory, Lexicase, etc. (e.g. Hudson, 1984; Mel'čuk, 1988; Starosta, 1988; Eroms, 2000). Thus far, there are many representations of dependency grammar. **Figure 1** displays two typical dependency representations of one sample sentence *We walked along the lake*.

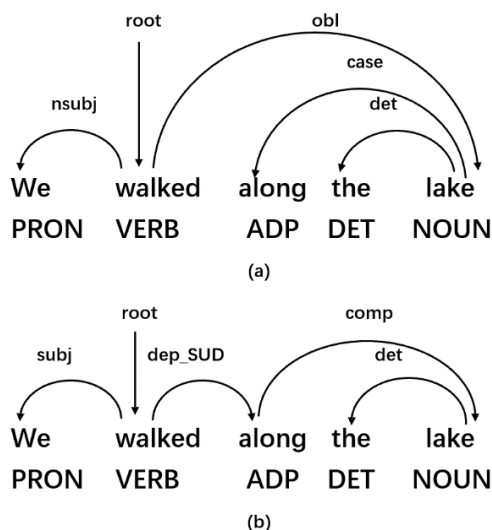


Figure 1. Dependency Representations of One English Sentence *We walked along the lake* Based on UD and SUD Annotation Schemes.

The dependency representation based on the Universal Dependencies (UD¹), as shown in **Figure 1 (a)**, is one of the most eminent models by now under the framework of dependency grammar. It attempts at establishing a multilingual morphosyntactic scheme to annotate various languages in a consistent manner (Nivre, 2015; Osborne and Gerdes, 2019). Thus, the UD annotation scheme holds a semantic over

¹See also <http://universaldependencies.org/>.

syntactic criteria to put priorities to content words to maximize “crosslinguistic parallelism” (Nivre, 2015; de Marneffe and Nivre, 2019). On the contrary, the Surface-Syntactic Universal Dependencies (SUD²) annotation scheme, as shown in **Figure 1 (b)**, follows the syntactic criteria to define not only the dependency labels but also the dependency links. It aims to make the annotation scheme close to the dependency traditions, like Meaning-Text-Theory (MTT) (Mel’čuk, 1988), Word Grammar (Hudson, 1984), etc. Hence, the SUD annotation scheme is a syntactic-oriented dependency representation that seeks to promote the syntactic motivations (Gerdes et al., 2018; Osborne and Gerdes, 2019). Therefore, the UD and SUD annotation schemes signify two typical preferences of dependency grammar, one is semantic-oriented, and the other is syntactic-oriented.

As shown in **Figure 1**, the linear sentence in both representations can be divided into several words; and the labelled arcs, directed from the governors to the dependents, represent different dependency types indicating the syntactic relations between elements within the sentence. Hence, the dependency representations indicate both the functional role of each word as well as the syntactic relations between different elements. More importantly, based on dependency representations, linguists have proposed several measurements for linguistic analysis. For one thing, **dependency distance** is defined as the linear distance of the governor and the dependent (Hudson, 1995). For another, the linear order of the governor and the dependent of each dependency type is referred to as **dependency direction** (Liu, 2010). When a governor appears before a dependent, the dependency direction is governor-initial or negative. Otherwise, it is governor-final or positive. For instance, in **Figure 1 (a)**, the arc above the dependent *we* and the governor *walked* forms a governor-final relation; and the dependency distance between these two elements is $2 - 1 = 1$ (the number 2 and 1 in the subtraction represent the linear order of the governor and dependent, respectively). Detailed calculating method will be shown in **Section 2**. Therefore, the dependency representations and the measures of dependency relations are both explicit and clear-cut. This explains the reason why dependency treebanks, i.e. corpora with annotations (Abeillé, 2003), are widespread among linguists in big-data era. As a result, the variations and universals of human languages are explored and unveiled through statistical and mathematical tools (Hudson, 1995; Liu et al., 2017). What is noteworthy is that previous studies have shown that **dependency distance** is an important indicator in demonstrating the notion of syntactic complexity and cognitive demand (Hudson, 1995; Gibson, 2000; Liu, 2008).

Under the framework of dependency grammar, Hudson (1995) characterized the definition of dependency distance based on the theories of memory decaying and short-term memory (e.g. Brown, 1958; Levy et al., 2013). The notion of syntactic difficulty and cognitive demand have been subsequently related to the linear distance between the governors and the dependents in cognitive science (Gibson, 1998; Hawkins, 2004). Based on a Romanian dependency treebank, Ferrer-i-Cancho (2004) hypothesized and proved that the mean distance of a sentence is minimized and constrained. These paved the way for Liu’s (2008) empirical study on dependency distance which provides a viable treebank-based approach towards the metric of syntactic complexity and cognitive constraint. Afterwards, series of studies exploring the relationship between dependency distance and syntactic and cognitive benchmarks have been conducted (e.g. Jiang and Liu, 2015; Wang and Liu, 2017; Liu et al., 2017). These studies share some similarities. First, it is well-corroborated that the frequency of dependency distance decreases with the increase of the dependency distance, viz., the distribution of dependency distance follows the linguistic law of the Least Effort Principle (LEP) or Dependency Distance Minimization (DDM) (Zipf, 1965; Liu et al., 2017). Second, it is believed that the greater the dependency distance is, the more difficult the sentence structure (Gibson, 1998; Hiranuma, 1999; Liu et al., 2017). Thus, the arithmetic average of all dependency distances of one sentence or a treebank or the **mean dependency distances** (MDDs) (Liu, 2008) has been an important index of memory burden, demonstrating the syntactic complexity and cognitive demand of the language concerned (Hudson, 1995; Liu et al., 2017).

Previous studies have shown that there are several factors that have effects on the measurement of dependency distance of a sentence, including sentence length, genre, chunking, language type, grammar, annotation scheme and so forth (e.g. Jiang and Liu, 2015; Wang and Liu, 2017; Lu et al., 2016; Hiranuma, 1999; Liu and Xu, 2012; Gildea and Temperley, 2010). Most of these factors have been well-investigated, however, the factor of annotation scheme has rarely been studied. Liu et al. (2009), for instance, investigated Chinese syntactic and typological properties based on five different Chinese

²See also <https://gitlab.inria.fr/grew/SUD>.

treebanks with different genres and annotation schemes, yet the treebanks adopted with different annotation schemes were used to avoid the corpus influences to ensure a reliable conclusion. Hence, the question as to the effects of annotation scheme on the distribution of dependency distance and MDD remains open.

Moreover, investigations into the benchmark of syntactic complexity and cognitive demand introduced above were primarily based on traditionally syntactic-oriented dependency models, for instance, the Stanford Typed Dependencies annotation scheme (de Marneffe and Manning, 2008) or other annotation schemes that specifically designed for each individual language. Thus, there is no consistency among different treebanks. In addition, although there are some qualitative investigations on the distinctions between the UD annotation scheme and various traditional syntactic-oriented annotation schemes (e.g. Osborne and Maxwell, 2015), and the existing studies also include some empirical studies focusing primarily on the consistently annotated UD scheme (e.g. Chen and Gerdes, 2017; 2018), it is still of our interest that, compared with those based on consistently annotated traditionally syntactic-oriented schemes, whether linguistic analysis based on the UD annotation scheme can still function as a metric of syntactic difficulty and cognitive demand, and if it can, what are the reasons for these distinctions?

Therefore, the deficiency of investigations into annotation scheme of treebanks leads to the inquiry of current study. We attempt at making comparisons of dependency distances based on two different annotation schemes, UD and SUD. Aimed to address the issues mentioned above, the following questions are under discussion based on UD and SUD treebanks:

- (1) Will the probability distribution of dependency distances of natural texts change when they are based on different annotation schemes? Do they still follow the linguistic law of DDM?
- (2) Based on MDDs, which annotation scheme is more congruent for the measurement of syntactic complexity and cognitive demand?
- (3) Which dependency types account most for the distinctions between UD and SUD annotation schemes?

2 Materials and Methods

Taking English language as an example, we adopt the Georgetown University Multilayer Corpus (GUM) (Zeldes, 2017) in UD 2.2 and SUD 2.2 projects. Both versions of the treebank are consisted of seven genres, viz. *academic writing*, *biographies*, *fiction*, *interviews*, *news stories*, *travel guides* and *how-to guides*. Since the treebanks are balanced in term of genres, it would better demonstrate the general features of the probability distribution of dependency distance when we adopt different annotation schemes.

To measure the effectiveness of MDDs as a metric of syntactic difficulty and cognitive demand in a broad sense, the *testing* sets of 20 languages with two versions of annotations were drawn from the UD 2.2 and SUD 2.2 to form 20 corresponding treebanks. There 20 languages are *Arabic (ara)*, *Bulgarian (bul)*, *Catalan (cat)*, *Chinese (chi)*, *Czech (cze)*, *Danish (dan)*, *Dutch (dut)*, *Greek (ell)*, *English (eng)*, *Basque (eus)*, *German (ger)*, *Hungarian (hun)*, *Italian (ita)*, *Japanese (jpn)*, *Portuguese (por)*, *Romanian (rum)*, *Slovenian (slv)*, *Spanish(sp)*, *Swedish (swe)* and *Turkish (tur)*. These 20 treebank-pairs would help to demonstrate the features and distinctions of syntactic- and semantic-oriented annotation schemes in measuring syntactic complexity and cognitive constraint.

As for the calculation of dependency distance, we adopted Jiang and Liu' (2015) approach. Formally, let $W_1...W_i...W_n$ be a word string. For any dependency relation between the words W_x and W_y ($x \geq 1, y \leq n$), if W_x is a head and W_y is its dependent, then the dependency distance between them is defined as the difference $x - y$; by this measure, the dependency distance of adjacent words is 1.

The MDD of the entire sentence can be defined as:

$$\text{MDD (sentence)} = \frac{1}{n-1} \sum_{i=1}^{n-1} |DD_i| \quad (1)$$

In this formula, n is the number of words in the sentence and DD_i is the dependency distance of the i -th syntactic relation of the sentence. Usually in a sentence there is one word (the *root* verb) without a head, whose DD is defined as zero.

The MDD of a treebank can be defined as:

$$\text{MDD}(\text{treebank}) = \frac{1}{n-s} \sum_{i=1}^{n-s} |DD_i| \quad (2)$$

Here, n is the total number of words in the sample, s is the total number of sentences in the sample and DD_i is the dependency distance of the i -th syntactic link of the sample.

When it comes to the MDD for a specific type of dependency relation in a sample, the formula can be shown as follows:

$$\text{MDD}(\text{dependency type}) = \frac{1}{n} \sum_{i=1}^n DD_i \quad (3)$$

In this case, n is the number of examples of that relation in the sample. DD_i is the dependency distance of the i -th dependency type.

For both UD and SUD annotations, the formats of their representations are CoNLL-X (de Marneffe & Manning, 2008). **Table 1** is a simplified CoNLL-X version of the sample sentence with UD annotation scheme.

Order	Word	Dependent			Head Order	Relation Dependency Type
		Lemma	POS	Feature		
1	<i>We</i>	we	PRON	PRP	2	nsubj
2	<i>walked</i>	walk	VERB	VBP	0	root
3	<i>along</i>	along	ADP	IN	5	case
4	<i>the</i>	the	DET	DT	5	det
5	<i>lake</i>	lake	NOUN	NN	2	obl

Table 1. Simplified Annotation of *We walked along the lake* in UD Treebank.

Take the first line in **Table 1** for example. It shows that the second word *walked* in the sentence has a dependent *we*, which is the first word of the sentence. The type of this dependency is *nsubj*, or *nominal subject*. As for the second line, it indicates that the *root* of the sentence is *walked*, signifying the head of the whole sentence rather than demonstrating a dependency relation; hence it is removed during computation. Regarding the sample sentence above, the DD of *nsubj* (line one) is $2 - 1 = 1$; *case* (line three) is $5 - 3 = 2$; *det* (line four) is $5 - 4 = 1$; *obl* (line five) is $2 - 5 = -3$. Hence, following formula (1), the MDD of the sentence can be obtained as follows: $(|1|+|2|+|1|+|-3|)/4=1.75$. Similarly, the MDD of the sample sentence based on SUD annotation scheme in **Figure 1 (b)** is $(|1|+|-1|+|+1|+|-2|)/4=1.25$.

3 Results and Discussion

Taking English language as an example, we would first focus on the probability distribution of dependency distance to investigate whether it follows the linguistic law of DDM when we adopt two distinctive annotation schemes. Following what Liu (2008) did, we would then calculate MDDs of 20 languages based on two annotation schemes to demonstrate which annotation is more effective to measure syntactic difficulty and cognitive demand. Finally, specific dependency types in the treebank of GUM would be under investigation to display the possible underlying explanation beneath the distinctions between these two annotation schemes.

3.1 Annotation Scheme and Probability Distribution of Dependency Distance

It is believed that dependency distance is cognitively restrained by human working memory (Liu et al., 2017). Therefore, human beings tend to minimize the dependency distances while interpreting or producing languages. Hence, based on different syntactic-oriented annotation schemes, it has been found that the probability distribution of dependency distances of natural languages follows similar distributional patterns, including right truncated zeta (Jiang and Liu, 2015; Wang and Liu, 2017; Liu et al., 2017) and right truncated waring (Jiang and Liu, 2015; Lu and Liu, 2016; Wang and Liu, 2017).

Following these researches, we fitted dependency distances of all 95 texts of GUM to these two probability distributions by the fitting program of probability distributions, Altmann-Fitter³. Since the determination coefficient R^2 can indicate the goodness-of-fit (Wang and Liu, 2017; Wang and Yan, 2018), the mean values of the determination coefficient R^2 in all seven genres were calculated. Conventionally, the excellent, good, acceptable and not acceptable goodness-of-fit for determination coefficient R^2 are 0.90, 0.80, 0.75 and less than 0.75, respectively.

It was found that the mean determination coefficient R^2 in the model fitting of right truncated waring and right truncated zeta based on both UD and SUD are larger than 0.80, indicating that the fitting results are good. In other words, the frequencies of dependency distances based on both UD and SUD treebanks can well capture the models of right truncated waring and right truncated zeta with a good coefficients of determination R^2 .

To conclude, the probability distributions of dependency distances of natural texts based on both UD and SUD annotation schemes share similar power law distribution, viz. the frequency of dependency distance decreases with the increase of the dependency distance. The results reveal that dependency distance distributions of all texts based on both UD and SUD follow the same regularity, supporting the Least Effort Principle (LEP) (Zipf, 1965) or the linguistic law of DDM (Liu, 2008; Futrell et al., 2015; Liu et al., 2017).

3.2 Annotation Scheme and Mean Dependency Distance

Except the probability distribution of dependency distance, the syntactic and cognitive parameter of MDDs is also of our interest. Hence, the MDDs of all 20 corresponding treebanks based on UD and SUD were calculated in accordance with formula (2). Our results show that although these two annotation schemes are divergent from each other, what is in common is that the MDDs of all 20 languages based on both annotation schemes are within 4 (Cowan, 2001), showing that the syntactic complexity of human languages is constrained by human cognitive limitation or LEP rather than annotation scheme itself. This is consistent with what we've discovered in **Section 3.1**. Moreover, what is noteworthy is that MDDs based on 20 UD treebanks are always larger than those based on SUD for each individual language. This means that language materials based on UD annotation scheme lead to the interpretation of larger MDDs. Theoretically, it is believed that annotation schemes that lead to shorter MDDs is more linguistically applicable (Osborne and Gerdes, 2019). Hence, the SUD annotation scheme seems to be more suitable for reflecting the human cognitive demand and the syntactic complexity of the language under processing.

This was followed by a dependent-samples t test. The result shows that MDDs based on UD ($M = 2.86$, $SD = .32$) are significantly longer than MDDs that based on SUD ($M = 2.52$, $SD = .39$), $t(19) = 11.10$, $p < .05$, $d^f = 2.48$. The p -value of .000 is less than .05, the null hypothesis that the means of MDD based on different annotation schemes are equal is rejected. Moreover, according to Cohen's conventions, the effect size of 2.48 corresponds to a large effect in practice and indicates that the MDDs based on UD was rated 2.48 standard deviations longer in distance than was SUD. Hence, with the distinction of annotation schemes, natural language texts based on UD annotation scheme tend to have longer MDD than that based on SUD.

In addition, the results of MDDs based on SUD are closer to those of Liu (2008: 174). This might be due to the fact that both the SUD annotation scheme and what Liu (2008) based on belong to the category of syntactic-oriented annotation schemes. Although some languages have larger MDDs (e.g. *Hungarian (hun)* and *Chinese (chi)*) and the other have smaller ones (e.g. *Turkish (tur)* and *Japanese (jpn)*) in Liu (2008) than those based on SUD, this might be attributed to that the annotation schemes of Liu (2008) are not consistently annotated across languages. Hence, when it comes to the relationship between annotation scheme and MDD, although still within a threshold of 4, MDD of language materials based on UD annotation scheme tends to be longer than that based on SUD, and the difference is significant. Moreover, the MDDs based on SUD share great similarities with those based on Liu (2008). Thus, to some extent, it can be summarized that the syntactic-oriented SUD is comparatively more expedient annotation scheme to researches concerning syntactic complexity and cognitive demand.

³See also <https://www.ram-verlag.eu/software-neu/software/>.

⁴A commonly used effect size statistic for the dependent-samples t test is d . In accordance with Cohen's (1988) conventions, small, medium, and large effect sizes for the dependent samples t test are .20, .50, and .80, respectively.

3.3 Annotation Scheme and Annotating Preference

In **Section 3.1** and **Section 3.2**, we investigated the universal inclination of DDM for natural languages and MDD as an indicator of syntactic complexity as well as cognitive demand based on two different annotation schemes. In **Section 3.3**, the reasons for the similarities as well as distinctions are of our interest. Since it is impossible to make detailed analysis based on all 20 corresponding treebanks, the English treebank of GUM is under investigation as a representative.

Since the SUD annotation scheme is near-isomorphic to the UD initiative (Gerdes et al., 2018), treebanks based on UD and SUD are very similar to a large extent. The greatest difference between UD and SUD treebanks is the direction of the dependency types used to indicate the relations between function words and content words. In this case, UD’s *aux*, *cop*, *mark* and *case* dependencies indicate dependency relations pointed from content words to function words (e.g. the *case* relation between *lake* and *along* as shown in **Figure 1 (a)**), while their directions are inverted in SUD and renamed as *comp* as shown in **Table 2** (e.g. the *comp* relation between *along* and *lake* in **Figure 1 (b)**) (Gerdes et al., 2018: 71). Meanwhile, other subordinate dependency relations remain intact.

UD Dependency	Corresponding SUD Dependency
<i>aux</i> , <i>cop</i> , <i>mark</i> , <i>case</i> , <i>xcomp</i> , <i>ccomp</i> , <i>obj</i> , <i>iobj</i> , <i>obl:arg</i>	<i>comp</i>

Table 2. General Corresponding Dependency Relations in UD and SUD Annotation Schemes.

As **Table 2** shows, the *comp* relation in SUD is consisted of more than four UD types (i.e. *aux*, *cop*, *mark* and *case*). Hence, according to Gerdes et al. (2018: 72), we nailed down the actual 4 corresponding pairs that differentiate the UD and SUD annotation schemes in **Table 3**. They are *aux* and *comp:aux*, *aux:pass* and *comp:pass*, *cop* and *comp:cop*, and finally, *mark & case* and *comp*.

UD		SUD	
Type	Relation	Type	Relation
<i>aux</i>	auxiliary	<i>comp:aux</i>	complement: auxiliary
<i>aux:pass</i>	passive auxiliary	<i>comp:pass</i>	complement: passive auxiliary
<i>cop</i>	copula	<i>comp:cop</i>	complement: copula
<i>mark</i>	marker	<i>comp</i>	complement: subordinating conjunction
<i>case</i>	case marking		complement: adposition

Table 3. Detailed Corresponding Dependency Relations in UD and SUD Annotation Schemes.

In **Table 3**, all the dependency relations in UD are head-final, and in SUD head-initial. In other words, nearly all *comp:aux*, *comp:pass*, *comp:cop* and *comp* in SUD designate the function words as heads over content words, hence the dependency directions are negative. Correspondingly, nearly all the *aux*, *aux:pass*, *cop* and *mark & case* in UD choose content words as head; hence the dependency directions are also altered. This shows that the underlying reason for the distinction between UD and SUD annotation scheme is that the UD annotation scheme favours taking the content words as the head of function words while the SUD annotation scheme chooses the function words as heads over content words in dependency relations (Nivre, 2015; Gerdes et al., 2018; Osborne and Gerdes, 2019). To be specific, the UD treebanks first connect the content words and then the function words to emphasize the semantic similarities of all languages, while the SUD treebanks connect content words mediated by function words to complete the functional roles of function words. Consequently, the more the number of modifiers before noun (between the apposition and the noun) is, the longer the dependency distance between the verb (*root*) and the noun. For instance, the distance between *walked* and *lake* in **Figure 1 (a)** would

enlarge if there are more modifier before the noun *lake*. Hence, the longer MDDs in UD treebanks can be attributed to the emphasis on semantic relations within sentence structure.

In fact, designating the head of linguistic structure has always been a focus of modern grammar, not only for dependency grammar but also for constituency-based frameworks (Jackendoff, 1977; Zwicky, 1985; Pollard and Sag, 1994), especially when it comes to the function words within sentence structure (de Marneffe and Nivre, 2019). The design of the SUD representation that prioritizes function words as heads over content words is in line with most traditional syntaxes (Hudson, 1984; Mel'čuk, 1988; Starosta, 1988; Eroms, 2000). Moreover, under the framework of dependency grammar, the status of function words has also been discussed by many theoretical studies (e.g. Groß and Osborne, 2015; Osborne and Maxwell, 2015). However, most studies focus on one aspect of function words or emphasize the qualitative features of dependency relations that related to function words. The current section provides some empirical evidence of the status of function words in semantic-oriented UD and syntactic-oriented SUD treebanks.

4 Conclusions and Implications

Through observation and calculation, it can be found that based on UD and SUD annotation schemes, natural English texts exhibit similar probability distribution. No matter what the genre of the text is, both share a power law distribution with a trend of minimizing dependency distance. This is also consistent with the well-exemplified DDM theory corroborated by Liu et al. (2017), showing the limitation of human working memory capacity.

When 20 corresponding treebanks are under investigation, the MDDs of Liu (2008)'s study and our study based on SUD annotation scheme are similar with each other, and they are significantly shorter than those based on UD, showing the consistency of syntactic-oriented annotation schemes and the possibility of applying SUD to language materials to measure syntactic complexity and cognitive demand.

Moreover, the reason underlying for the distinctions between UD and SUD annotation schemes is the dependency types indicating the relations between apposition and noun. The UD annotation scheme prefers a semantic orientation, while the SUD annotation scheme favours a syntactic orientation which holds a function-word priority. To be specific, the corresponding pairs, *aux* and *comp:aux*, *aux:pass* and *comp:pass*, *cop* and *comp:cop*, and *mark & case* and *comp* in UD and SUD annotation schemes lead to longer MDDs of UD treebanks.

Therefore, the current study suggests that, to some extent, the consistently syntactic-oriented annotation scheme (SUD) is better than the consistently semantic-oriented one (UD) in linguistic analysis of syntactic complexity and human cognitive demand. However, it is still worthwhile to spare more efforts to assess the effectiveness of consistently annotated syntactic-oriented representation to capture both the variations and universals of natural human languages.

References

- Amir Zeldes. 2017. The GUM corpus: creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3):581-612. DOI: 10.1007/s10579-016-9343-x
- Anne Abeillé. 2003. *Treebank: Building and Using Parsed Corpora*. Kluwer Academic Publisher, Dordrecht.
- Arnold M. Zwicky. 1985. Heads. *Journal of Linguistics*, 21:1-29.
- Carl J. Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.
- Daniel Gildea and David Temperley. 2010. Do grammars minimize dependency length? *Cognitive Science*, 34(2): 286-310.
- Edward Gibson. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68(1):1-76. DOI: 10.1016/s0010-0277(98)00034-1
- George K. Zipf. 1965. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Hafner Publishing Company, New York.
- Haitao Liu. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2):159-191.

- Haitao Liu. 2009. *Dependency Grammar: From Theory to Practice*. Science Press, Beijing.
- Haitao Liu. 2010. Dependency direction as a means of word-order typology a method based on dependency treebanks. *Lingua*, 120(6):1567-1578.
- Haitao Liu and Chunshan Xu. 2012. Quantitative typological analysis of Romance languages. *Poznań Studies in Contemporary Linguistics*, 48(4):597-625.
- Haitao Liu, Chunshan Xu, and Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, 21:171-193.
- Haitao Liu, Yiyi Zhao, and Wenwen Li. 2009. Chinese syntactic and typological properties based on dependency syntactic Treebanks. *Poznań Studies in Contemporary Linguistics*, 45(4):509-523.
- Hans Jürgen Heringer. 1993. Dependency syntax-basic ideas and the classical model. *Syntax-An International Handbook of Contemporary Research*, volume 1, 298-316.
- Hans-Werner Eroms. 2000. *Syntax der deutschen Sprache*. Walter de Gruyter, Berlin.
- Igor Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. The SUNY Press, Albany, N.Y.
- Jingyang Jiang and Haitao Liu. 2015. The effects of sentence length on dependency distance, dependency direction and the implications—Based on a parallel English–Chinese dependency Treebank. *Language Sciences*, 50:93-104.
- Joakim Nivre. 2015. Towards a Universal Grammar for Natural Language Processing. *Computational Linguistics and Intelligent Text Processing*, 3-16. DOI: 10.1007/978-3-319-18111-0_1
- Joel E. Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Johanna Nichols. 1986. Head-marking and dependent-marking grammar. *Language*, 62:56-119.
- John A. Hawkins. 2004. *Efficiency and Complexity in Grammars*. Oxford University Press, Oxford.
- John Brown. 1958. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10:173-189.
- Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. 2018. SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. In *Proceedings of Universal Dependencies Workshop 2018*, 66-74. Brussels.
- Lucien Tesnière. 1959. *Éléments de syntaxe structurale*. Klincksieck, Paris.
- Marie-Catherine de Marneffe and Christopher D. Manning. 2008. Stanford Typed Dependencies Manual. *Technical Report*, 338-345.
- Marie-Catherine de Marneffe and Joakim Nivre. 2019. Dependency Grammar. *Annual Review of Linguistics*, 5:197-218.
- Nelson Cowan. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87-185.
- Peter H. Matthews. 1981. *Syntax*. Cambridge University Press, Cambridge.
- Qian Lu, Chunshan Xu, and Haitao Liu. 2016. Can chunking reduce syntactic complexity of natural languages? *Complexity*, 21(S2):33-41.
- Ramon Ferrer-i-Cancho. 2004. Euclidean distance between syntactically linked words. *Physical Review E*, 70:056135.
- Ray S. Jackendoff. 1972. *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge.
- Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *PNAS*, 112:10336-10341.
- Richard Hudson. 1984. *Word Grammar*. Basil Blackwell, New York.
- Richard Hudson. 1995. Measuring syntactic difficulty. <http://dickhudson.com/wp-content/uploads/2013/07/Difficulty.pdf>
- Roger Levy, Evelina Fedorenko, and Edward Gibson. 2013. The syntactic complexity of Russian relative clauses. *Journal of Memory and Language*, 69 (4):461-495.

- So Hiranuma. 1999. Syntactic difficulty in English and Japanese: A textual study. *UCL Working Papers in Linguistics*, 11:309-322.
- Stanley Starosta. 1988. *The Case for Lexicase: An Outline of Lexicase Grammatical Theory*. Pinter Publishers, New York.
- Thomas Groß and Timothy Osborne. 2015. The dependency status of function words: auxiliaries. In Eva Hajičová and Joakim Nivre (eds.), *Proceedings of the 3rd International Conference on Dependency Linguistics*, pp. 111-120. Stroudsburg, PA: Assoc. Comput. Linguist.
- Timothy Osborne and Daniel Maxwell. 2015. A historical overview of the status of function words in dependency grammar. In Eva Hajičová and Joakim Nivre (eds.), *Proceedings of the 3rd International Conference on Dependency Linguistics*, pp. 241-250. Stroudsburg, PA: Assoc. Comput. Linguist.
- Timothy Osborne and Kim Gerdes. 2019. The status of function words in dependency grammar: A critique of Universal Dependencies (UD). *Glossa: A Journal of General Linguistics*, 4(1):17.1-28. DOI: 10.5334/gjgl.537
- Yaqin Wang and Haitao Liu. 2017. The effects of genre on dependency distance and dependency direction. *Language Sciences*, 59:135-147.
- Yaqin Wang and Jianwei Yan. 2018. A quantitative analysis on a literary genre *Essay*'s syntactic features. In Jingyang Jiang & Haitao Liu (eds.), *Quantitative Analysis of Dependency Structures*, pp. 295-314. Berlin/Boston: De Gruyter Mouton.