Mean Hierarchical Distance
Augmenting Mean Dependency Distance

Yingqi Jing
Department of Linguistics
Zhejiang University
Hangzhou, China
`jakenevergivesup@gmail.com`

Haitao Liu
Department of Linguistics
Zhejiang University
Hangzhou, China
`htliu@163.com`

Abstract

With a dependency grammar, this study provides a unified method for calculating the syntactic complexity in linear and hierarchical dimensions. Two metrics, mean dependency distance (MDD) and mean hierarchical distance (MHD), one for each dimension, are adopted. Some results from the Czech-English dependency treebank are revealed: (1) Positive asymmetries in the distributions of the two metrics are observed in English and Czech, which indicates both languages prefer the minimalization of structural complexity in each dimension. (2) There are significantly positive correlations between sentence length (SL), MDD, and MHD. For longer sentences, English prefers to increase the MDD, while Czech tends to enhance the MHD. (3) A trade-off relationship of syntactic complexity in two dimensions is shown between the two languages. English tends to reduce the complexity of production in the hierarchical dimension, whereas Czech prefers to lessen the processing load in the linear dimension. (4) The threshold of the MDD_2 and MHD_2 in English and Czech is 4.

1 Introduction

The syntactic structures of human languages are generally described as two-dimensional, and many structural linguists use tree diagrams to represent them. For example, Tesnière (1959) employed tree-like dependency diagrams called *stemmas* to depict the structure of sentences. Tesnière also distinguished between linear order and structural order. In this study, we follow Tesnière's clear-cut separation of these two dimensions and investigate the relation between them by using an English and Czech dependency treebank, designing different measures to quantify the complexity of syntactic structure in each dimension.

The relationship between linear order and structural order is a crucial topic for all structural syntax. For Tesnière (1959: 19), structural order (hierarchical order) preceded linear order in the mind of a speaker. Speaking a language involves transforming structural order to linear order, whereas understanding a language involves transforming linear order to structural order. It is worth mentioning that Tesnière's stemmas do not reflect actual word order, but rather they convey only hierarchical order. This separation of the two ordering dimensions has had great influence on the development of dependency grammar and word-order typology. The ability to separate the two dimensions has been argued to be an advantage for dependency grammar, since it is more capable than constituency grammar of examining each dimension independently (Osborne, 2014).

The real connection between hierarchical order and word order is evident when the principle of projectivity or continuity is defined in dependency grammar (see, e.g., Lecerf, 1960; Hays, 1964: 519; Robinson, 1970: 260; Mel'čuk, 1988: 35; Nivre, 2006: 71). According to Hudson (1984: 98),

> "if A depends on B, and some other element C intervenes between them (in linear order of strings), then C depends directly on A or on B or on some other intervening element."

Projectivity is immediately visible in dependency trees; a projective tree, as shown in Figure 1, has no crossing lines. But it must be mentioned that projectivity is not a property of the dependency tree in itself, but only in relation to the linear string of words (Nivre, 2003: 51), and some languages with relatively free word order (e.g., German, Russian, and Czech) have more crossing lines than languages with relatively rigid word order (Liu, 2010: 1576). Here, we also use the term "pro-

jection" in linear algebra as a means of transforming a two-dimensional syntactic structure to one-dimensionality. Thus, in a projective or nonprojective dependency tree, the string of words is just an image projected by the structural sentence onto the spoken chain, which extends successively on a timeline.
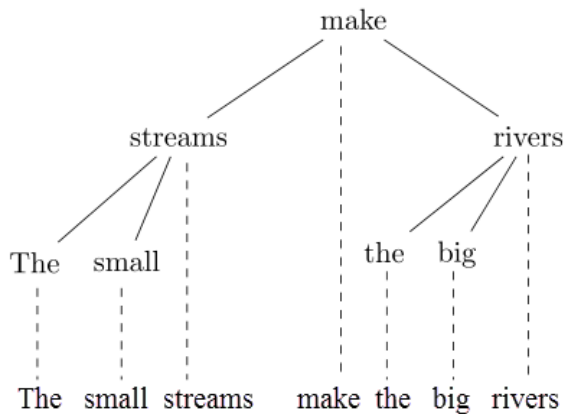

Figure 1: A dependency tree of *The small streams make the big rivers*.[1]

This study focuses on exploring the structural rules of English and Czech using two metrics, mean dependency distance (MDD), as first explored by Liu (2008), and mean hierarchical distance (MHD), as introduced and employed here for the first time. These metrics help predict language comprehension and production complexity in each dimension. The metrics are mainly based on the empirical findings in psycholinguistics and cognitive science, and we tend to bind the two dimensions of syntactic structure together. To assess the value of these metrics, we have explored the syntactic complexity of English and Czech with the help of the Prague Czech-English Dependency Treebank 2.0 (PCEDT 2.0).

The rest of this manuscript introduces the PCEDT 2.0 and data pre-processing in Section 2. The theoretical background and previous empirical studies concerned with the two metrics (MDD and MHD) are presented in Section 3, and our methods for calculating them are also given in this section. In Section 4, we present the results and findings, which are summarized in the last section.

2   Czech-English dependency treebank

The material used in this study is the PCEDT 2.0, which is a manually parsed Czech-English parallel corpus, sized at over 1.2 million running words in almost 50,000 sentences for each language (Hajič et al., 2012). The English part of the PCEDT 2.0 contains the entire Penn Treebank-Wall Street Journal (WSJ) Section (Linguistic Data Consortium, 1999). The Czech part consists of Czech translations of all of the Penn Treebank-WSJ texts. The corpus is 1:1 sentence-aligned. The parallel sentences of both languages are automatically morphologically annotated and parsed into surface-syntax dependency trees according to the Prague Dependency Treebank 2.0 (PDT 2.0) annotation scheme. This scheme acknowledges an analytical layer (a-layer, surface syntax) and a tectogrammatical layer (t-layer, deep syntax) of the corpus (Hajič et al., 2012). Only the a-layer was used for the current study. More information about the treebank and its annotation scheme is available on the PCEDT 2.0 website.[2]
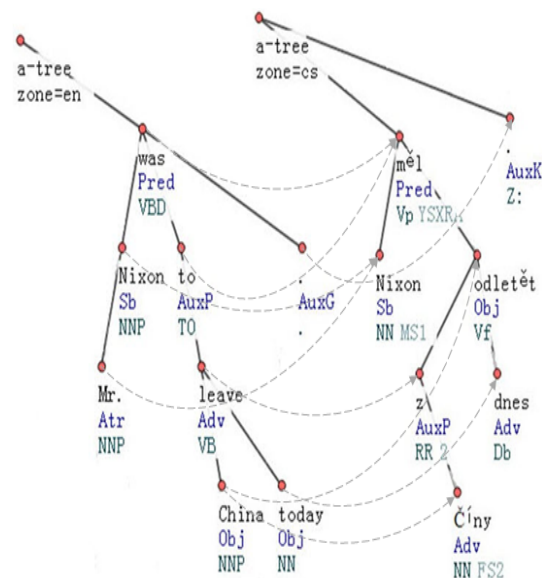

Figure 2: A sample parallel sentence at the a-layer

PCEDT 2.0 is a strictly aligned corpus, which is stored as *.treex format using the XML-based Prague Markup Language (PML). It can be easily visualized with the tree editor TrEd and displayed as the sample parallel sentence (en. *Mr. Nixon was to leave China today.* cs. *Nixon měl z Číny odletět dnes.*) in Figure 2. The word alignment is indicated by the dashed grey arrows pointing from

---

[1]The sentence *The small streams make the big rivers* is the English translation of Tesnière's (1959: 19) example, but linear order and projection lines have been added to the stemma.

the English part to the Czech part.

We first extract data from the original Treex documents with R 3.0.2, supported by the XML package for parsing each node of the treebank, and restore it into a Microsoft Access database. The transformed corpus is much easier to access and analyze (Liu, 2009: 113). Table 1 shows a previous English sample sentence converted into a new format, and the header contains sentence number (sn), word number (wn), word (w), part-of-speech (POS), governor number (gn), governor (g) and dependency relations (dep). The root verb is the only word that has no governor and we indicate its lack of a governor and governor number using 0.

sn	wn	w	POS	gn	g	dep
1770	1	Mr.	NNP	2	Nixon	Atr
1770	2	Nixon	NNP	3	was	Sb
1770	3	was	VBD	0	0	Pred
1770	4	to	TO	3	was	AuxP
1770	5	leave	VB	4	to	Adv
1770	6	China	NNP	5	leave	Obj
1770	7	today	NN	5	leave	Obj
1770	8	.	.	3	was	AuxG

Table 1: A converted sample sentence in English

The a-layer of the corpus contains 1,173,766 English nodes and 1,172,626 Czech word tokens, which are combined into 49,208 parallel sentences. Sentences with less than three words (e.g., @, *Virginia:*, *New Jersey:*) or some special four-element sentences (e.g., *"Shocked."*, *Právníci jistě ne.*) were removed from each language (477 and 474 sentences). They are mainly specific markers in the news or incomplete sentences. Finally, the intersection of two language sets constitutes the corpus used in our study according to the sentence number. Table 2 presents an overview of our corpus with 48,647 parallel sentences (s), and the mean sentence length (msl) of English and Czech is 24.1 and 23.63, respectively. However, Czech has a much higher percentage of non-projective (n.p.) dependencies than English.

name	size	s	msl	n.p.
en	1172244	48647	24.1	0.01%
cs	1149630	48647	23.63	3.11%

Table 2: General description of the corpus

3 Mean dependency distance and mean hierarchical distance

Previous scholars have devoted a lot of effort to building a well-suited metric for measuring and predicting syntactic complexity of all human languages, for instance, Yngve's (1960; 1996) Depth Hypothesis[3] and Hawkins' (2003; 2009) principle of Domain Minimalization. The current psycholinguistics and cognitive science have also provided evidence for this issue. Gibson (1998; 2000) conducted many reading experiments and proposed a Dependency Locality Theory (DLT), which associates the increasing structural integration cost with the distance of attachment. Fiebach et al. (2002) and Phillips et al. (2005) observed a sustained negativity in the ERP signal during sentence regions with filler-gap dependencies, indicating increased syntactic integration cost. These studies have a common interest in connecting linear dependency distance with language processing difficulty.

The concept of "dependency distance (DD)" was first put forward by Heringer et al. (1980: 187) and defined by Hudson (1995: 16) as "the distance between words and their parents, measured in terms of intervening words." With the previous theoretical and empirical evidence, Liu (2008: 170) proposed the mean dependency distance (MDD) as a metric for language comprehension difficulty and gave the formula in (1) to calculate it.

$$MDD = \frac{1}{n} \sum_{i=1}^{n} |DD_i| \qquad (1)$$

In this formula, n represents the total number of dependency pairs in a sentence, and $|DD_i|$ is the absolute value of the i-th dependency distance. It must be noted that DD can be positive or negative, denoting the relative position or dependency direction between a dependent and its governor. Thus, the MDD of a sentence is the average value of all pairs of $|DD_i|$.

The present study builds on this distance-based notion of dependencies and extends the concept into the hierarchical dimension. The act of listening involves transforming a linear sentence

---

[3]Yngve took a constituency-based view and measured the depth of a sentence by counting the maximum number of symbols stored in the temporary memory when building a syntactic tree. Yngve's model and mertic are specifically designed for sentence production.

into a two-dimensional syntactic tree; this bottom-up process is concerned with integrating each linguistic element with its governor and forms a binary syntactic unit. Storage or processing costs occur when a node has to be retained in the listener's working memory before it forms a dependency with its governor (Gibson, 1998). This theory has laid the fundations of many comprehension-oriented metrics.

Conversely, the act of speaking involves transforming a stratified tree to a horizontal line. This top-down process is almost like a spreading activation where the activation of a concept will spread to neighboring nodes (Hudson, 2010: 74-79). Then each concept can be expressed and pronounced sequentially on a timeline. The complexity of this activation procedure is hypothesized and measured by the conceptual distance between the root of a sentence and some other nodes.

The major evidence supporting our assumption is the empirical findings of code-switching by Eppler (2010; 2011), and Wang and Liu (2013). They report that the MDD of mixed dependencies (words from distinct languages) is larger than that of monolingual ones, suggesting that increased processing complexity can actually promote code-switching. These conclusions are drawn from the studies on German-English and Chinese-English code-switching. However, Eppler, and Wang and Liu have only concentrated on investigating the phenomena from the listener's perspective in terms of MDD; they neglect the fact that one of the major motivations for code-switching is to lessen a speaker's production load.[4] For instance, appropriate words or phrases are not instantly accessible, so the speaker seeks some alternative expressions in another language to guarantee continuity in speech. This trade-off relation may provide a starting point to measure the structural complexity from the speaker's perspective.

A stratified syntactic tree can be projected horizontally, and we record the relative distance between each node and the root, as shown in Figure 3. Non-projective sentences can be represented in the same way. Here, we take the root of a syntactic tree as a reference point and designate its projection position as 0; it is the central node

and provides critical information about syntactic constituency (Boland et al., 1990; Trueswell et al., 1993). The vertical distance between a node and the root, or the path length traveling from the root to a certain node along the dependency edges, is defined as "hierarchical distance (HD)". For example, the HD of the word *China* in Figure 3 is 3, which denotes the vertical distance or path length between the node and the root.
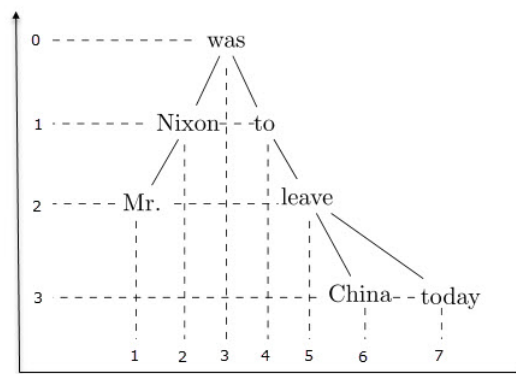

Figure 3: Projection of a dependency tree in two dimensions

The average value of all HDs in a sentence is the mean hierarchical distance (MHD). In this study we hypothesize that the MHD is a metric for predicating the structural complexity in the hierarchical dimension. It can be expressed with formula (2).

$$MHD = \frac{1}{n}\sum_{i=1}^{n} HD_i \qquad (2)$$

According to the formulas (1) and (2), we can calculate MDD and MHD of the sample sentence in Figure 3. The MDD of this sentence is (1+1+1+1+1+2)/6=1.17 and the MHD is (2+1+1+2+3+3)/6=2. Note that punctuation marks are rejected when measuring the MDD and MHD.

Furthermore, these two metrics can be applied to measure a text or treebank. To do this, one need merely average the MDD and the MHD of all the sentences in the text or treebank, and in so doing the results represent the MDD and the MHD of the language at hand. In the following parts, we use MDD_2 and MHD_2 to represent the measures at the textual level. For a text with a specific number of sentences (s), its MDD_2 and MHD_2 can be calculated with (3) and (4), respectively.

---

[4]Some scholars may focus on the social motivations of code-switching, such as accommodating oneself to a social group, but the present study tends to emphasize its psychological property.

$$MDD_2 = \frac{1}{s} \sum_{j=1}^{s} MDD_j \qquad (3)$$

$$MHD_2 = \frac{1}{s} \sum_{j=1}^{s} MHD_j \qquad (4)$$

To sum up, the syntactic structure of language has two dimensions, which can be reduced to one dimension by means of orthogonal projections. Two statistical metrics (MDD and MHD), one for each dimension, are proposed. These metrics measure syntactic complexity. To be more specific, MDD is actually a comprehension-oriented metric that measures the difficulty of transforming linear sequences into layered trees, whereas MHD is a production-oriented metric that measures the complexity of transforming hierarchical structures to strings of words. These metrics are applicable at both the sentential and the textual levels. In the next section, we further investigate the relations and distributions of MDD and MHD in English and Czech sentences.

4 Results

Section 3 defined the two metrics, MDD and MHD, and gave their corresponding formulas for calculation. In this section, we first calculate the MDD and MHD of each sentence in English and Czech, and describe their distributions in nature. The correlations between sentence length (SL), MDD, and MHD are then tested. Further, we extend the two metrics to the textual level, and compare the MDD_2 and MHD_2 of English and Czech. Finally, the threshold of the two metrics in both languages is investigated.

4.1 Asymmetric distributions of MDD and MHD

Hawkins (2003: 122; 2009: 54) proposed a Performance-Grammar Correspondence Hypothesis (PGCH),

> "grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments".

The PGCH predicts an underlying correlation between variation data in performance and the fixed conventions of grammars. In other words, the more preferred a structure X is, the more productively grammaticalized it will be, and the easier it is to process due to the frequency effect (Harley, 1995: 146-148; Hudson, 2010: 193-197).

The patterns of syntactic variation can reflect the underlying processing efficiency; hence we first focus on describing the distributions of MDD and MHD of each sentence in the treebank. Figure 4 exhibits two positively skewed distributions of MDD and MHD when the SL (no punctuations) of each English sentence equals 10. The Pearson's moment skewness coefficients (Sk) are 1.31 and 0.78.[5] The coefficients indicate that most English sentences with 10 words get MDD and MHD values below the mean.
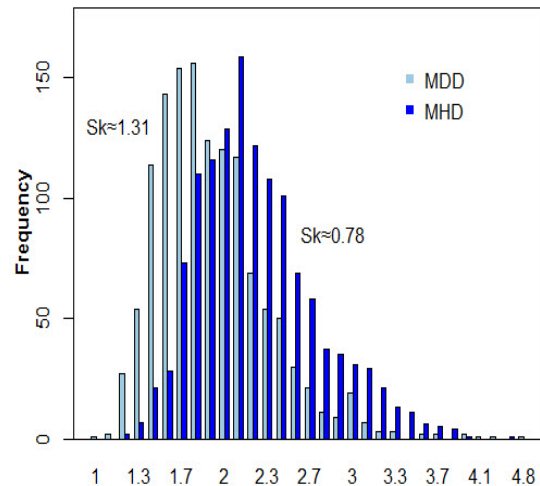

Figure 4: Asymmetric distributions of MDD and MHD for English sentences (SL=10)

Some other types of English and Czech sentences of different lengths, the frequency of which is more than 50 times in the treebank, are also positively skewed in the distribution of MDD and MHD, as shown in Figure 5. The skewness coefficients of the two metrics of both languages are all positive, fluctuating around 1, though there is no significant correlation between SL and Sk. It appears that the mass of both English and Czech sentences, of whatever length, tend to have lower

---

[5]The Pearson's moment coefficient of skewness is measured by the formula $(Sk = \mu_3 / \mu_2^{\frac{3}{2}})$, where μ_2 and μ_3 are the second and third central moments. For a symmetric distribution, if the data set looks the same to the left and right of the center point, the skewness value is equal to zero. If $Sk > 0$, it is a positive skewing indicating more than half of the data below the mean, whereas if $Sk < 0$, it is negatively skewed with more data above the mean.
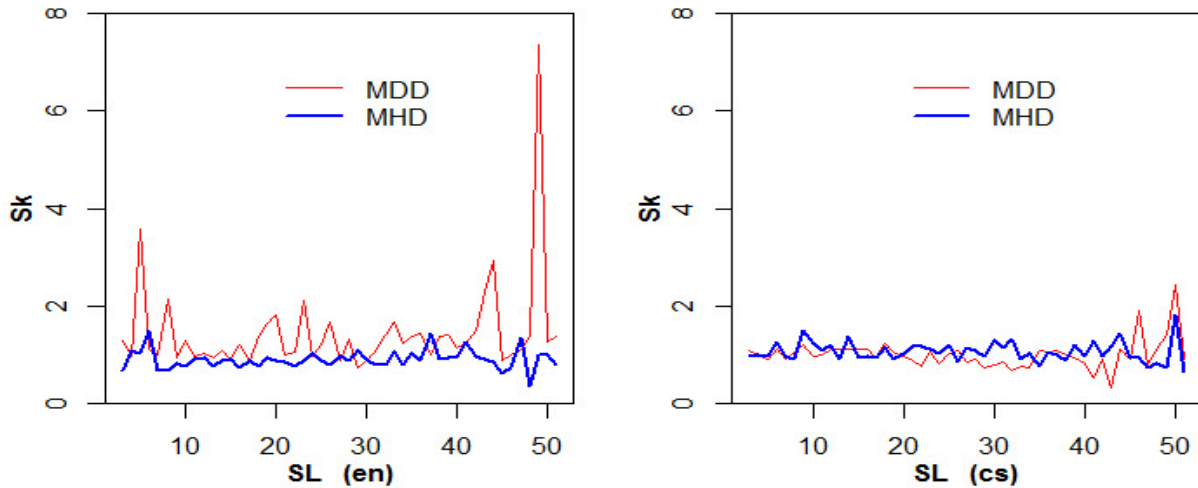
Figure 5: Relationships between SL and Sk in MDD and MHD

MDD and MHD values. Why are lower MDD and MHD preferred in both languages? If grammars are assumed to be independent of processing (Chomsky, 1969), no such consistent asymmetric distributions of the two metrics in different language types would be expected. One possibility for accounting for the skewness is that syntactic rules are direct responses to processing ease and are grammaticalizations of efficiency principles (Hawkins, 1994: 321). Hence, we can observe these preferences in two dimensions, and both English and Czech tend to minimize the MDD and MHD values. The minimalization of these two metrics reflects the efficiency principle of human language.

4.2 Correlations between SL, MDD, and MHD

Another relevant issue concerning the MDD and MHD is whether these metrics can predict the structural complexity for varying sentence lengths in different languages. Table 3 displays the positive correlations between SL, MDD and MHD in English and Czech, and they are all significantly correlated ($p<0.01$). Correlation coefficients (Cor) between SL and MHD in English and Czech are the highest (0.74 and 0.74, respectively), which is followed by moderate correlations (0.54 and 0.42) between SL and MDD in the two languages. The MDD and MHD in both languages are the least correlated with each other, but they are also significant.

More precisely, we build a linear regression model to fit the data. The goodness of fit (R^2) and slope (k) can be used to evaluate the model and predict the increase rate of the two languages. The R^2 between SL and MHD is acceptable at 0.54 and 0.54, while the other two pairs in each language get pretty low values. The slope of the SL-MHD fitting line in English (0.09) is slightly lower than that in Czech (0.12), which suggests the increase of SL will bring more gains of MHD in Czech than in English.

We also visualize the relationships between MDD and MHD of English and Czech sentences with a scatter plot in Figure 6. Although a large overlap is shown between MDD and MHD, we can still observe different extensions in each language. If the SL is taken as a moderator variable, English sentences tend to increase the MDD for longer sentences, whereas Czech sentences prefer higher MHD as the SL is increasing. This variation of preference in different languages can also be predicted by the above linear model. From the perspective of language processing, English sentences prefer to enhance the comprehension difficulty rather than the production cost as the sen-

Lang	X-Y	Cor	p	k	R^2
	SL-MDD	0.54	<0.01	0.03	0.3
en	SL-MHD	0.74	<0.01	0.09	0.54
	MDD-MHD	0.19	<0.01	0.41	0.04
	SL-MDD	0.42	<0.01	0.02	0.18
cs	SL-MHD	0.74	<0.01	0.12	0.54
	MDD-MHD	0.11	<0.01	0.36	0.01

Table 3: Correlations between SL, MDD, and MHD

tences get longer; on the contrary, Czech sentences prefer increasing the structural complexity in hierarchical dimension, which is assumed to be connected with the production load here.
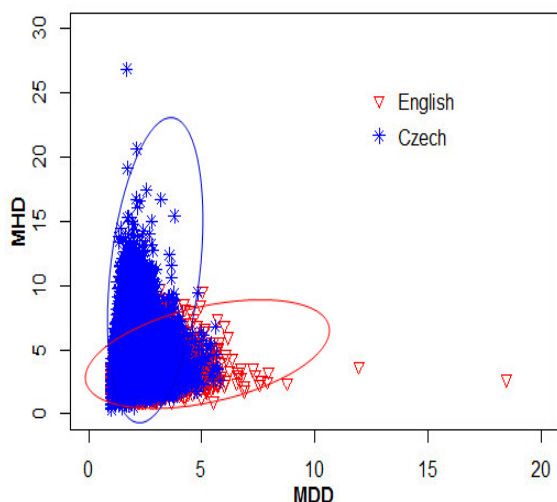

Figure 6: Relationships between MDD and MHD of English and Czech sentences

4.3 Trade-off relation between MDD_2 and MHD_2

The two metrics can be expanded to measure the MDD_2 and MHD_2 of certain languages as well, and compare the values across different language types. English and Czech are both mitigated languages with a subject-verb-object (SVO) word order, but the word order of Czech is relatively unrestricted, whereas English word order has been claimed to become rigid due to the loss of case inflections (Tesnière, 1959: 33; Vennemann, 1974; Steele, 1978; Liu, 2010). Due to this high degree of word order variation, it is almost inevitable for Czech to have more non-projective structures than English. Will the high percentage of non-projective dependency relations in Czech enlarge its MDD_2, or will the two metrics even differentiate the syntactic complexity across the two languages?

Figure 7 represents the MDD_2 and MHD_2 of English and Czech. The MDD_2 of English is 2.31 and that of Czech is 2.18. These numbers are similar to Liu's (2008) results, which were arrived at by investigating the MDD_2 of twenty languages. The MHD_2 is 3.41 for English and 3.78 for Czech. All values are below 4. English and Czech both get a lower MDD_2 than MHD_2, but the MDD_2 of Czech is slightly lower than that of English, even

though Czech has a much higher percentage of non-projectivity. Projectivity is of course widely viewed as a constraint in natural language parsing, but the number of projectivity violations that actually occur does not appear to have predictive value for language processing difficulty in the linear dimension.

There seems to be a zero-sum property of the two metrics in different languages. English gains a relatively higher MDD_2 than Czech but has a lower MHD_2. Conversely, even though the MDD_2 of Czech is not as high as that of English, its MHD_2 is greater than that of English. This reciprocal relationship is given at the sentential level in Figure 6, and is also shown at the textual level in Figure 7. This trade-off relation between the structural complexity in the two dimensions partially proves the dynamic balance of code-switching from the listener's and speaker's perspectives.

This also reveals that the weights of the two metrics are not equal in varying language types. English tends to reduce the structural complexity in the hierarchical dimension, while Czech prefers to lessen the processing cost in the linear dimension.
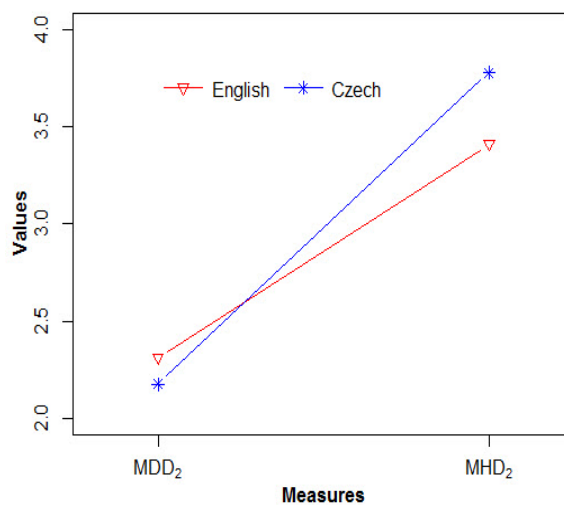

Figure 7: MDD_2 and MHD_2 of English and Czech

4.4 Threshold of MDD_2 and MHD_2

The two metrics, MDD_2 and MHD_2, can differentiate the syntactic complexity or difficulty between English and Czech in each dimension. But can they reveal any common attribute between varying languages? Cowan (2001) claimed that a more precise capacity limit of short-term memory should be about four chunks on the average, and Liu (2008) also observed a threshold of MDD_2
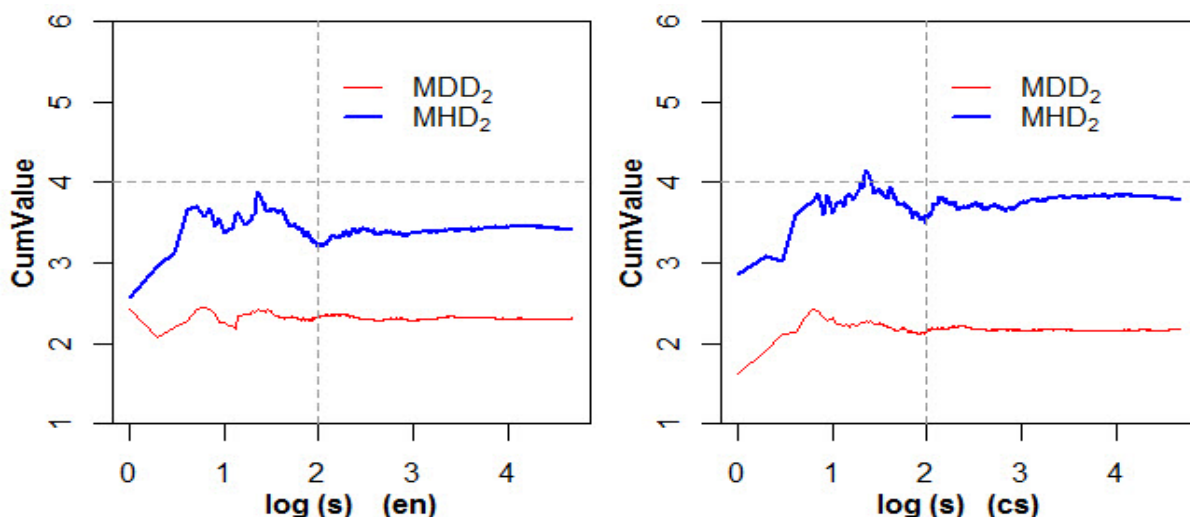
Figure 8: Cumulative average values of MDD_2 and MHD_2 in English and Czech

for twenty languages at about 4. Does there exist a universal boundary value in the hierarchical dimension?

To answer these questions, we make a time-series plot to characterize real-time variation of MDD_2 and MHD_2 in English and Czech, as shown in Figure 8. Due to a large quantity of sentences, the horizontal axis of the plot is scaled logarithmically. A high degree of variation in MDD_2 and MHD_2 is displayed at first, and when more sentences (about 10^2 sentences) are added in, the cumulative average values become stable in both languages. In this plot, we can also find that the maximum values of MDD_2 and MHD_2 in the two languages are below 4,[6] though a small part of the MHD_2 value in Czech is above 4. This minor deviation is mainly caused by fewer sentences and some extreme examples. It should be noted that the corpus used in the present study has a relatively long mean sentence length (around 24 words per sentence), and some sentences with fewer words are also removed, which will, to some extent, enlarge the MDD_2 and MHD_2 of the two languages. But a threshold of the MDD_2 and MHD_2 below 4 is shown as well, and we believe that there do exist boundary conditions for syntactic structure in the two dimensions, and the threshold is largely due to the capacity limits of short-term memory.

Thus, the capacity limit of working memory can be described in the process of both language comprehension and production, and a similar boundary value of 4 reflects their internal coherence.

5 Conclusions

We have presented a systematic study of how to measure the complexity of the syntactic structures of human languages, extending previous distance-based theories. Two statistical metrics (MDD and MHD) have been proposed for predicting the structural complexity of language, one for each dimension. The MDD is comprehension-oriented by measuring the difficulty of speaking, whereas the MHD is production-oriented, calculating the cost of listening. The two metrics are applicable at both the sentential and the textual levels.

Data from the Czech-English dependency treebank have been used to test and justify our approach. Some major findings are summarized as follows. (1) Positive asymmetries in the distributions of the MDD and MHD are observed in English and Czech. Both languages prefer to minimize the processing ease in each dimension. (2) There are significantly positive correlations between SL, MDD, and MHD. For longer sentences, English prefers to increase the MDD, while Czech tends to enhance the MHD. (3) A reciprocal relationship of syntactic complexity in the two dimensions is shown between English and Czech, which indicates an imbalance in weight of MDD_2 and MHD_2. English tends to reduce the syntactic complexity in the hierarchical dimension, whereas Czech prefers to lessen the processing load in the linear dimension. (4) The threshold of MDD_2 and MHD_2 in the two languages is 4 (even below 3 for the MDD_2), which suggests internal coherence for

---

[6]The MDD_2 for English and Czech is even below 3, but for another language in Liu's (2008) study, i.e. Chinese, the MDD_2 was 3.66.

the process of language comprehension and production.

More quantitative work is needed for the two metrics, especially concerning empirical validity in the arena of psycholinguistics. Furthermore, typological studies are another potentially useful direction for exploration.

Acknowledgments

References

Julie E. Boland, Michael K. Tanenhaus, and Susan M. Garnsey. 1990. Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language*, 29(4): 413–432.

Noam Chomsky. 1969. *Aspects of the Theory of Syntax*. MIT press.

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.

Eva Eppler. 2010. *Emigranto: the syntax of German-English code-switching*. Vienna: Braumüller.

Eva Eppler. 2011. The Dependency Distance Hypothesis for bilingual code-switching. In *Proceedings of the International Conference on Dependency Linguistics*, pages 145–154. Barcelona, Spain, 5–7 September.

Christian J. Fiebach, Matthias Schlesewsky, and Angela D. Friederici. 2002. Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47(2): 250–272.

Edward Gibson. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68(1): 1–76.

Edward Gibson. 2000. The dependency locality theory: a distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita, and Wayne O'Neil (eds.), *Image, language, brain: papers from the First Mind Articulation Project Symposium*, pages 95–126. Cambridge, MA: MIT Press.

Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and

Zdeněk Žabokrtský. 2012. Prague Czech-English Dependency Treebank 2.0 LDC2012T08. DVD. Philadelphia: Linguistic Data Consortium.

Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. 2012. Announcing Prague Czech-English Dependency Treebank 2.0. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, pages 3153–3160. Istanbul, Turkey, 21–27, May. European Language Resources Association (ELRA).

Linguistic Data Consortium. 1999. Penn Treebank 3. LDC99T42.

Trevor Harley. 1995. *The Psychology of Language*. Hove: Psychology Press.

John Hawkins. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

John Hawkins. 2003. Efficiency and complexity in grammars: three general principles. In John C. Moore, and Maria Polinsky (eds.), *The Nature of Explanation in Linguistic Theory*, pages 121–152. Stanford, Calif: CSLI Publications.

John Hawkins. 2009. Language universals and the performance-grammar correspondence hypothesis. In Morten H. Christiansen, Chris Collins, and Shimon Edelman (eds.), *Language Universals*, pages 54–78. Oxford: Oxford University Press.

David Hays. 1964. Dependency theory: a formalism and some observations. *Language*, 40(4): 511–525.

Hans-Jürgen Heringer, Bruno Strecker, and Rainer Wimmer. 1980. *Syntax. Fragen - Lösungen - Alternative*. München: Fink.

Richard Hudson. 1984. *Word Grammar*. Oxford: Blackwell.

Richard Hudson. 1995. Measuring syntactic difficulty. Unpublished paper. *URL www.phon.ucl.ac.uk/home/dick/papers/difficulty.htm*

Richard Hudson. 2010. *An Introduction to Word Grammar*. Cambridge: Cambridge University Press.

Yves Lecerf. 1960. Programme des Conflits, Modèle des Conflits. *Traduction Automatique*, 1(4): 11–18; 1(5): 17–36.

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*. Beijing: Science Press.

Haitao Liu. 2010. Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua*, 120(6): 1567–1578.

Igor Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. Albany, NY: State University of New York Press.

Joakim Nivre. 2003. An Efficient Algorithm for Projective Dependency Parsing. In *Proceedings of the 8th International Workshop on Parsing Technologies (IWPT 03)*, pages 149–160. Nancy, France, 23–25 April.

Joakim Nivre. 2006. *Inductive Dependency Parsing*. Netherlands: Springer.

Timothy Osborne. 2014. Dependency grammar. In Andrew Carnie, Yosuke Sato, and Daniel Siddiqi (eds.), *The Routledge Handbook of Syntax*, pages 604–626. London: Routledge.

Colin Phillips, Nina Kazanina, and Shani H. Abada. 2005. ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22(3), 407–428.

Jane Robinson. 1970. Dependency structures and transformational rules. *Language*, 46(2): 259–285.

Susan Steele. 1978. Word order variation: A typological study. In Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), *Universals of human language, vol. 4: Syntax*, pages 585–624. Stanford: Stanford University Press.

Lucien Tesnière. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck.

John C. Trueswell, Michael K., and Christopher Kello. 1993. Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(3), 528–553.

Theo Vennemann. 1974. Topics, subjects and word order: From SXV to SVX via TVX. In John M Anderson, and Charles Jones (eds.), *Historical linguistics*, pages 339–376. Amsterdam: North-Holland Pub. Co.

Lin Wang and Haitao Liu. 2013. Syntactic variations in Chinese–English code-switching. *Lingua*, 123: 58-73.

Victor Yngve. 1960. A model and an hypothesis for language structure. In *Proceedings of the American philosophical society*, pages 444–466. 17, October. Philadelphia: American Philosophical Society.

Victor Yngve. 1996. *From grammar to science: New foundations for general linguistics*. Amsterdam & Philadelphia: John Benjamins.