

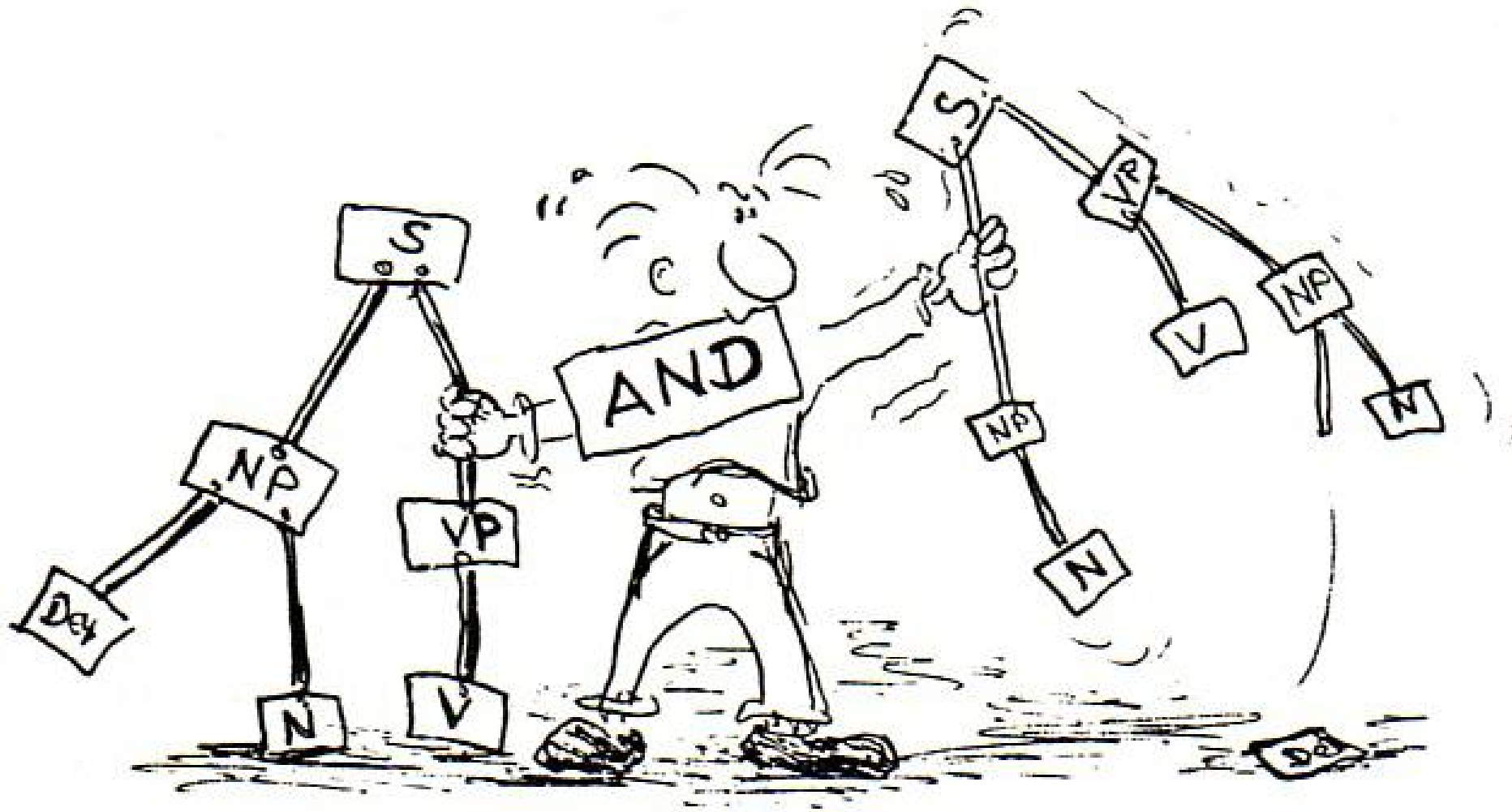


***A syntactic complexity analysis of
translational English text
across genres***

Jiajin Xu

Beijing Foreign Studies University

Transplanted trees (COTE) versus trees in their homeland (FLOB)





Overview of talk

- **Corpus approach to Translation Universals**
- **The predicament of current TU research**
- **Methodological innovations**
- **A syntactic complexity analysis of translational English text across genres**



Corpus-based studies of TUs

- Translation Universals (Baker 1993, 1996)
- Not without challenges
- TUs are innovative in that translation is seen as a legitimate variety of language.
- Translationese, hybrid language, the third code, etc.



Features of translation

- The test of the hypotheses
- Characteristic features of translation
- ***Explication***
- ***Simplification***
- ***Normalisation***
- ***Levelling out, etc.***



Descriptive translation research

- A handful of surface features of lexis have been examined in search of the universal translational traits (Blum-Kulka 1986; Toury 1995; Mauranen & Kujamäki 2004).



Over- and under-representation

- **Word forms:** e.g. *hapax legomena*, *connectives*, *foreign words*, *unusual collocation*, *(standardised) type-token ratio*.
- **Word classes:** e.g. *pronouns*, *nouns*, *content words*, *function words*, or the ratios based on these.
- Some **length measures**, e.g. mean word/sentence length.



The predicament of current TU research

- Syntactic features of translational language have been largely sidestepped, if not overlooked, in corpus-based translation studies.
- Only a limited number of grammatical elements, e.g., sentence length, passive voice and verb tense, have been occasionally dealt with.



Methodological innovations

- Computational advances in NLP
- A number of parsers are available for syntactic annotation and analysis.
- The current parsers are less accurate or reliable than PoS taggers.
- Good enough for grammatical category based research.



A case study of syntactic complexity

- My study of the syntactic complexity of translational text takes into account **fourteen measures** (cf. Lu 2010; Xu & Jia 2011) based on the counting of sentences, clauses, T-units, dependent clauses, coordinate phrases, complex nominals, and verb phrases



The 14 S/C indices

- **MLC**: mean length of clauses; **MLS**: mean length of sentences; **MLT**: mean length of T-units; **C/S**: clauses per sentence; **C/T**: clauses per T-unit; **CT/T**: complex T-units per T-unit; **DC/C**: dependent clauses per clause; **DC/T**: dependent clauses per T-unit; **CP/C**: coordinate phrases per clause; **CP/T**: coordinate phrases per T-unit; **T/S**: T-units per sentence; **CN/C**: complex nominals per clause; **CN/T**: complex nominals per T-unit; **VP/T**: verb phrases per T-unit

(Lu 2010)



The 14 S/C indices

- **MLC**: mean length of clauses;
- **MLS**: mean length of sentences;
- **MLT**: mean length of T-units;
- **C/S**: clauses per sentence;
- **C/T**: clauses per T-unit;
- **CT/T**: complex T-units per T-unit;
- **DC/C**: dependent clauses per clause;



The 14 S/C indices

- **DC/T**: dependent clauses per T-unit;
- **CP/C**: coordinate phrases per clause;
- **CP/T**: coordinate phrases per T-unit;
- **T/S**: T-units per sentence;
- **CN/C**: complex nominals per clause;
- **CN/T**: complex nominals per T-unit;
- **VP/T**: verb phrases per T-unit



Dimensions of syntactic complexity

- The 14 syntactic complexity indices are categorised as (adapted from Lu 2010):
 - length measures
 - sentence complexity (?)
 - subordination
 - coordination
 - key constituent units



Corpora

- The datasets used in the study are made up of the **COTE** corpus (Corpus of Translational English, compiled by Richard Xiao) and **FLOB** corpus (compiled by Christian Mair).
- COTE and FLOB represent comparable translational and original English texts of early 1990s respectively.



Corpus profile

- The two corpora have the total word tokens of one million words respectively.
- 500 texts of translational English (COTE)
- 500 texts of native English (FLOB)



Four genres of the corpora

Genres	15 text types	500 texts
Press	A-C	88 texts
General prose	D-H	206 texts
Learned	J	80 texts
Fiction	K-R	126 texts



Parsing and S/C index retrieval

- (BFSU) Stanford Parser (Windows GUI)

(ROOT
(S
(NP
(NP (NNP Hezb) (NNPS Forces) (NNP Threaten) (NNS Attacks))
(PP (IN on)
(NP
(NP (NNP Airports) (NNP Hong) (NNP Kong) (NNP AFP))
(PP (IN in)
(NP (NNP English) (CD 0935) (NN GMT) (CD 10) (NNP Sep) (N
(VP (VBD warned)
(NP (NNP Thursday))
(SBAR
(S
(NP (PRP they))
(VP (MD would)
(VP (VB start)

(ROOT

(S

(NP

(NP (NNP Hezb) (NNPS Forces) (NNP Threaten) (NNS Attacks))

(PP (IN on)

(NP

(NP (NNP Airports) (NNP Hong) (NNP Kong) (NNP AFP))

(PP (IN in)

(NP (NNP English) (CD 0935) (NN GMT) (CD 10) (NNP Sep) (N

(VP (VBD warned)

(NP (NNP Thursday))

(SBAR

(S

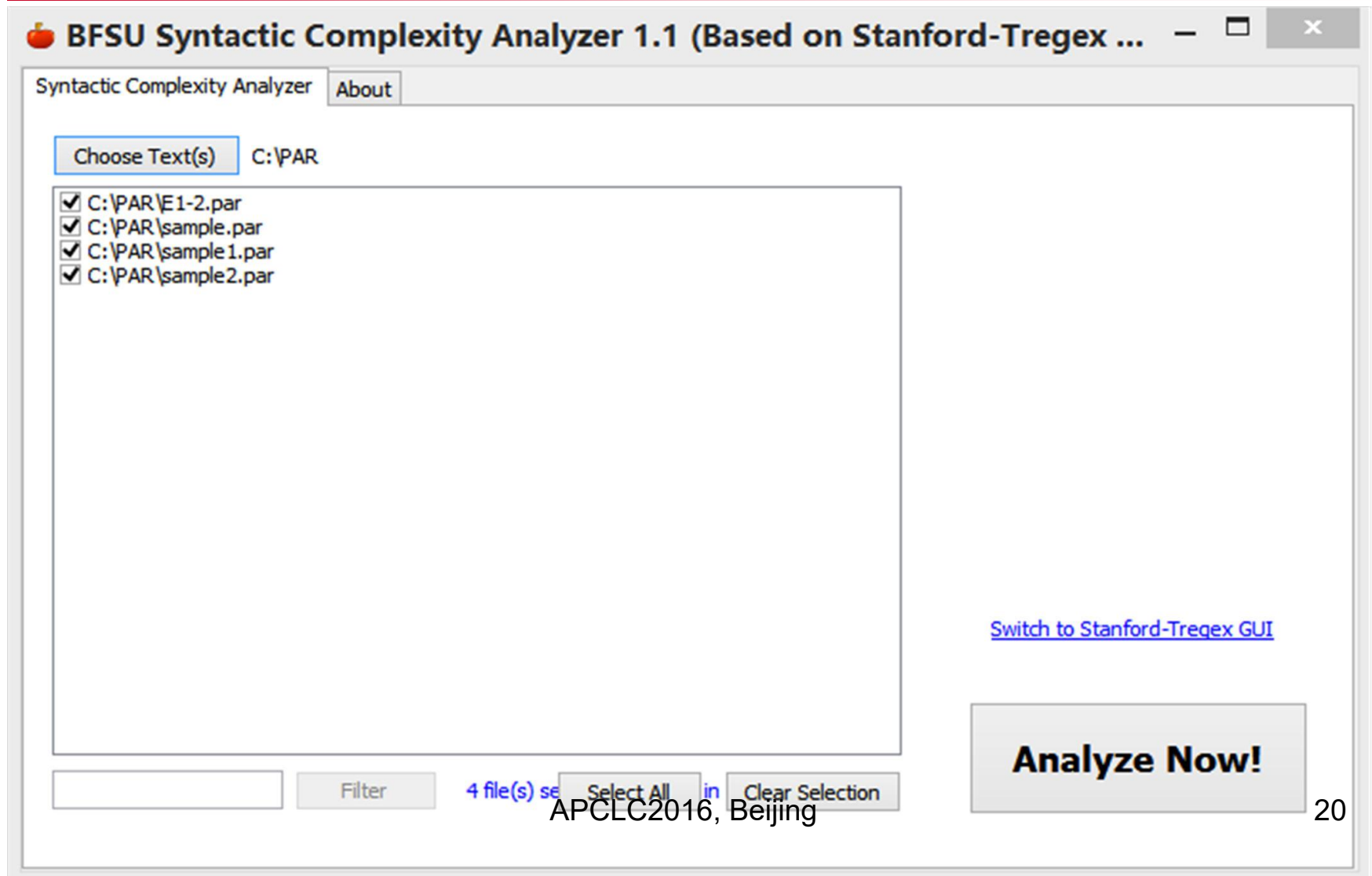
(NP (PRP they))

(VP (MD would)

(VP (VB start)



BFSU Syntactic Complexity Analyser





Results

COTE vs FLOB



Syntactic complexity b/t COTE & FLOB

Measure type	Code	t	Sig.
Length measures	MLC	3.28	.00
	MLS	5.82	.00
	MLT	5.87	.00
Sentence complexity	C/S	3.80	.00
Subordination	C/T	4.36	.00
	CT/T	3.00	.00
	DC/C	4.30	.00
	DC/T	4.70	.00
Coordination	CP/C	4.01	.00



COTE is more complex.

- In 13 out of 14 syntactic complexity indices, translational English is significantly more complex than native English.



However



Syntactic complexity comparison across genres

COTE-**Learned** and FLOB-
Learned



Syntactic complexity comparison between COTE-Learned and FLOB-Learned

Measure type	Code	t	Sig.
Length measures	MLC	-.182	.86
	MLS	2.44	.02
	MLT	1.65	.10
Sentence complexity	C/S	2.32	.02
Subordination	C/T	1.62	.11
	CT/T	1.11	.27
	DC/C	1.37	.17
	DC/T	1.47	.14
Coordination	CP/C	.46	.65



No difference!

- Almost no difference was found between translational scholarly English and native scholarly English texts in terms of syntactic complexity.



Likewise



Syntactic complexity comparison across genres

COTE-**General prose** and
FLOB-**General prose**

Syntactic complexity comparison between COTE-General prose & FLOB-General prose

Measure type	Code	t	Sig.
Length measures	MLC	1.03	.30
	MLS	.41	.68
	MLT	1.59	.11
Sentence complexity	C/S	1.02	.31
Subordination	C/T	1.78	.08
	CT/T	.34	.74
	DC/C	2.01	.05
	DC/T	2.33	.02
Coordination	CP/C	1.56	.12



No difference!!

- Almost no difference was found between translational English general prose and English general prose in terms of syntactic complexity, **EITHER**.



***Who on earth produced the
complex translational English texts?***



Syntactic complexity comparison across genres

COTE-**Press** and FLOB-**Press**



Syntactic complexity comparison between COTE-Press and FLOB-Press

Measure type	Code	t	Sig.
Length measures	MLC	10.10	.00
	MLS	8.83	.00
	MLT	11.20	.00
Sentence complexity	C/S	-.16	.88
Subordination	C/T	3.13	.00
	CT/T	1.75	.08
	DC/C	2.15	.03
	DC/T	2.81	.01
Coordination	CP/C	8.65	.00



- In 11 out of 14 syntactic complexity indices, translational English **news texts** is significantly more complex than native English.



Syntactic complexity comparison across genres

COTE-**Fiction** and FLOB-
Fiction



Syntactic complexity comparison between COTE-Fiction and FLOB-Fiction

Measure type	Code	t	Sig.
Length measures	MLC	4.89	.00
	MLS	5.69	.00
	MLT	5.43	.00
Sentence complexity	C/S	4.52	.00
Subordination	C/T	3.91	.00
	CT/T	3.93	.00
	DC/C	3.94	.00
	DC/T	4.22	.00
Coordination	CP/C	3.25	.00



- The translational English **fiction** texts are significantly more complex than native English **fiction** texts along all fourteen syntactic complexity measures.



Finding

- **Translational English fiction** is the main contributor of syntactic complexity of translational English.
- **Translational news texts** are another key player in making the translational English more complex syntactically.



Afterthoughts

- Literary text-based translation studies might be biased.
- Balanced corpus-based study and cross-genre comparison can present a fuller picture.



Afterthoughts

- Are longer sentences more complex?
- Longer coordinated structures do not necessarily add to the complexity of sentences.

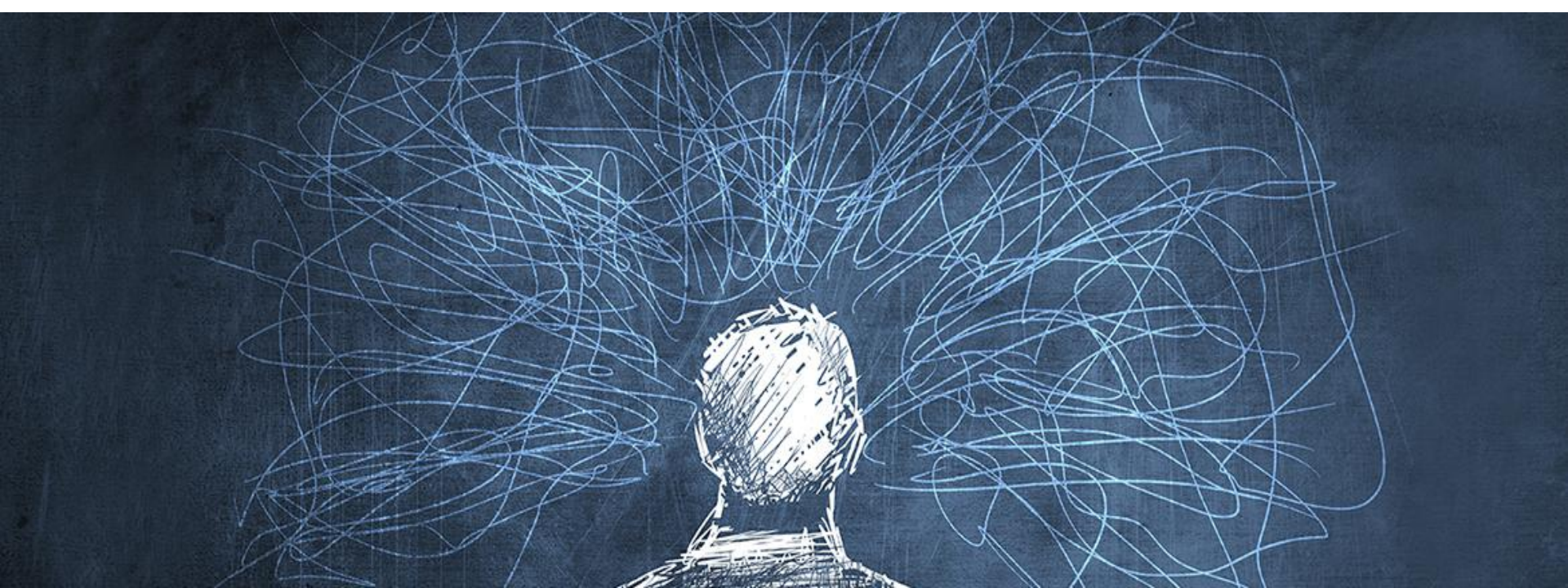


References

- Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.). *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins. 233-250.
- Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In Harold Somers (ed.). *Terminology, LSP and Translation*. Amsterdam: John Benjamins. 175-86.
- Blum-Kulka, Shoshana. 1986. Shifts of cohesion and coherence in translation. In Juliane House & Shoshana Blum-Kulka (eds.). *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Narr. 17-35.
- Lu, Xiaofei. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15(4): 474-496.
- Mauranen, Anna & Pekka Kujamäki (eds.). 2004. *Translation Universals: Do They Exist?* Amsterdam: John Benjamins.
- Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.
- Xu, Jiajin & Yunlong Jia. 2011. BFSU Syntactic Complexity Analyser. Beijing: National Research Centre for Foreign Language Education, Beijing Foreign Studies University.



Thank you!





What is a T-unit?

- a T-unit is a dominant clause and its dependent clauses: as Hunt said, it is "one main clause with all subordinate clauses attached to it" (Hunt 1965:20).
- A simple sentence is T-unit.
- A coordinate sentence is a T-unit.
- A main clause together with the subordinate clauses is a T-unit.



Syntactic complexity scale

- Simple sentence → coordinated sentence → subordinated sentences, complex nominals
1. I have a son. He is ten.
 2. I have a son and he is ten.
 3. I have a son who is ten.
 4. I have a ten-year-old son.

Wolfe-Quitero et al. (1998: 76)