Using comparable corpora for interpreters' terminology preparation

Ran Xu

ranxu2015@outlook.com

China Foreign Affairs University Beijing, China

Challenges

- Interpreters often work in a wide range of domains and have limited time to prepare for and activate domain-specific terminologies before interpreting.
- The terminological resources for technical meetings are rarely specific enough for interpreters to use straight away.
- Interpreters have to spend a lot of time reading through meeting documents. linterpreters' term lists are largely done manually.



Current situation for research

- Limited research into the use of modern term extraction and concordance tools for the task of SI.
- A few previous studies mentioned the application of corpora as potential tools for interpreters.

eg. Fantinuoli (2006), Gorjanc (2009), Rütten (2003)

 However, no empirical study to test whether the use of term extraction and concordance tools can help interpreters increase their preparation efficiency and to what extent the use of the tools influences interpreters' SI performance.

Research objectives

- To investigate how to integrate the use of corpus tools into interpreters' terminology preparation
- To measure the effect of using the proposed preparation procedure and the corpus tools on simultaneous interpreting performance.



The preparation procedure



Corpus-based preparation procedure and tools used



Tools to use

• Term extraction tool: Syllabs

В	C	D	E	F	G	H	I
EN	general_count	specific_count	general_frequency	specific_frequency	quotient_of_occurence	entity	form
se fuel	0	20	7.87E-08	0.000481486	6119.732707	NP	used fuel
w	0	19	7.87E-08	0.000458558	5828.316863	ABR	hlw
с	0	17	7.87E-08	0.000412702	5245.485177	ABR	jnc
rs	0	16	7.87E-08	0.000389774	4954.069334	ABR	fbrs
sile	0	13	7.87E-08	0.00032099	4079.821804	ADJ	fissile
fl	0	12	7.87E-08	0.000298063	3788.405961	ABR	jnfl
wrs	0	12	7.87E-08	0.000298063	3788.405961	ABR	pwrs
ps	0	10	7.87E-08	0.000252207	3205.574275	ABR	sfps
ontainment pressure	0	10	7.87E-08	0.000252207	3205.574275	NP	containment pressure
wrs	0	10	7.87E-08	0.000252207	3205.574275	ABR	bwrs
ontainment structure	0	9	7.87E-08	0.000229279	2914.158432	NP	containment structure
actor core	0	9	7.87E-08	0.000229279	2914.158432	NP	reactor core
olation condenser system	0	9	7.87E-08	0.000229279	2914.158432	NP	isolation condenser s
)S	0	9	7.87E-08	0.000229279	2914.158432	ABR	nps
ctinides	0	8	7.87E-08	0.000206351	2622.742589	NOUN	actinides
/h	0	8	7.87E-08	0.000206351	2622.742589	ABR	twh
drogen production	0	8	7.87E-08	0.000206351	2622.742589	NP	hydrogen production
r	0	8	7.87E-08	0.000206351	2622.742589	ABR	lwr
esign basis	0	8	7.87E-08	0.000206351	2622.742589	NP	design bases
op	1	16	1.57E-07	0.000389774	2477.034667	ABR	npp
perimental reactor	0	7	7.87E-08	0.000183423	2331.326745	NP	experimental reactor

Concordance tool: SketchEngine's concordance function

	Query reaction	sos (c,sos,s per mator)
n	Page 1	of 19 Go Next Last
	file1671096	Fast Breeder Reactor (FBR) power stations can generate electricity
	file1671096	promoting the development of fast breeder reactors as the major nuclear power source for the
	file1671096	Engineering Center and the Experimental Fast Reactor JOYO . In the same way that experience
	file1671096	Experimental FBR JOYO JOYO , Japan 's first fast reactor , started operation in 1978 . Initially
	file1671096	upgrade the irradiation test capacity of the reactor . This reactor serves as a research facility
	file1671096	irradiation test capacity of the reactor . This reactor serves as a research facility in the developme
IS	file1671096	was successfully recycled back into the reactor , thus completing a nuclear fuel cycle
s	file1671096	the commercialization of the fast breeder reactor . For detailed information , click the
	file1671096	description of JOYO , Japan 's first fast breeder reactor , and the important experimental work and
	file1671096	D Facilities Most of JNC 's fast breeder reactor R&D facilities are located at the O-arai
	file1671096	analysis program is being carried out using the reactor JOYO and overseas FBR facilities . In the
	file1671096	said that none of Japan's 50 idled nuclear reactors would restart until the NRA issued its
	file1671096	So who has the power to restart nuclear reactors in Japan ? My guess is that it is the NRA
	file1671096	it is the NRA approving the safety of the reactor and any restart plans , followed by the
	file1671096	power to approve the restart of nuclear reactors if they meet safety requirements (Japan
	file1671096	addition , Prime Minister Noda ordered two reactors in Oi , Fukui Prefecture , to be restarted
	file1671096	really cannot oppose restarting the nuclear reactors in light of the urgent need for power and
ос	file1671096	cost of both permanently shutting down all reactors (they have no spent fuel repository yet
	file1671096	to recover economically , they need these reactors restarted now . And as Economics Minister
	file1671096	called for the government to nationalize the reactors and to immediately begin decommissioning
	Page 1	of 19 Go Next Last

Node forms Doc IDs Collocations

Concordance

Word List Word Sketch Thesaurus Find X Sketch-Diff Corpus Info

Save as subcorp View option KWIC Sentence Sort Left Right Node Reference Shuffle Sample Filter Overlaps 1st hit in Frequency Node tags

This preparation procedure is expected to bring a twofold benefit to simultaneous interpreters:

- Helping to form and manage their tailor-made terminology resources in their work environments
- Boosting readiness of relevant terms for quick access and retrieval in SI.

Design of a pilot experiment



Experiment setting and procedure

	Group	Preparation method	Preparation Time
Pilot	Control	Traditional	9 days
Experiment (FR)	Test	Using auto-lists and concordancer	9 days



Dependent variables

- SI performance scores
- Terminological accuracy scores
- Terminology error types and error numbers
- Degrees of departures for each error type





The effect of using both auto-lists and the concordancer in interpreting preparation

The pilot experiment with 22 trainee interpreters shows:

- The test group had significantly higher term accuracy scores (个7.5%). (P<0.05)
- The test group made significantly fewer **term omission errors (OM)** in SI tasks (\downarrow 9.3%). (P<0.05)
- The test group had significantly better **post-task recall of terms** than the group without using any tool (\uparrow 18%). (P<0.05)
- The test group spent significantly less preparation time than the control group. (↓17%). (P>0.05)

- An increased level of term density in the source speech and working into B language affected trainee interpreters' performance by deteriorating term accuracy in interpretations.
- However, using both tools played a significant role in mitigating the detrimental effects of increased term density in the source speech and working into B language on the numbers of serious errors and omission errors (OM) in interpretations.

Summary of results

- The preparation procedure using both the term extractor and the concordancer within ample preparation time yielded the better preparation result then the traditional preparation procedure.
- Using both tools generally helped to improve trainee interpreters' terminological performance during SI by increasing term accuracy scores by 7.5% and reducing the number of omission errors by 9.3%.

 On the other hand, terminology preparation (through using both tools) is not a "magical cure" for all. The data shows that the preparation procedure only helped to improve students' holistic SI scores to a certain extent (but not yielding any statistical significance).

Pedagogical implications

- This study demonstrates that training on terminology preparation by using comparable corpora could be a useful supplement to the already existing professional interpreting training.
- It is important for both students and trainers to be aware that corpora and corpus tools when used properly can assist interpreters' terminology preparation and achieve an enhanced performance.



Limitation

 Providing the automatically-generated term lists to the test group was to avoid overburdening them. However, this is arranged at the expense of a weaker ecological validity of the experiment, as in a real interpreting assignment, the interpreters have to get the tem lists by using the tool themselves.



Future work

Future experiment:

- the participants use term extraction tool themselves
- SketchEngine platform



Thank you!

