Getting Across in Medical Communication:

A Corpus-based Approach to Analyse and Improve Comprehensibility

of Machine Translation

Yanmeng Liu	**Meng Ji**	**Pierrette Bouillon**
FASS/SLC	FASS/SLC	FTI/TIM
University of Sydney	University of Sydney	Université de Genève
yliu6746@uni.sydney.edu.au	christine.ji@university.edu.au	pierrette.bouillon@unige.ch

Appendices

Table 1. Multi-level bariables for the linguitic comaprision

Levels	Variables
Lexical	Type/token ratio (TTR)
	Word length (WLS)
	Lexical frequency profile (LFP)
Sentence	Sentence length (SLS)
	Frequency of Complex sentences (CSC)
	Frequency of Conjunction (CONJ)
Grammar Functional	Frequency of Preposition (PREP)
	Frequency of Modal words (MD)
	Frequency of Determiners (DTM)

Table 2. Examples of French source texts.

en cas d'hospitalisation, la couverture d'assurance est-elle en chambre commune?
avez-vous été vacciné contre la rougeole?
pouvez-vous énumérer ou écrire les noms de médicaments auxquels vous êtes allergiques?

Table 3. Comparison of MT and NBC64

Variables	MT	BNC64
TTR	0.1831	0.0143
WLS	3.2662	1.9567
SLS	11.1893	20.8870
CSC	0.0031	0.0032
CONJ	0.0031	0.02891
PREP	0.0579	0.0652
MD	0.0282	0.0105
DTM	0.0720	0.0673

Table 4 Lexical Frequency Profile

Word List	MT/%	BNC64/%
One	70.58	88.22
Two	3.60	2.81
Three	0.31	1.19
Not in the list	25.51	7.78

Table 5 Examples for layman expressions

Medical terms	Lay terms
cardiac	heart; involving the heart
dermatologic	of the skin
hepatitis	inflammation of the liver
hypertension	high blood pressure
oncology	the study of tumours or cancer
pulmonary	lung
respiratory	breathing