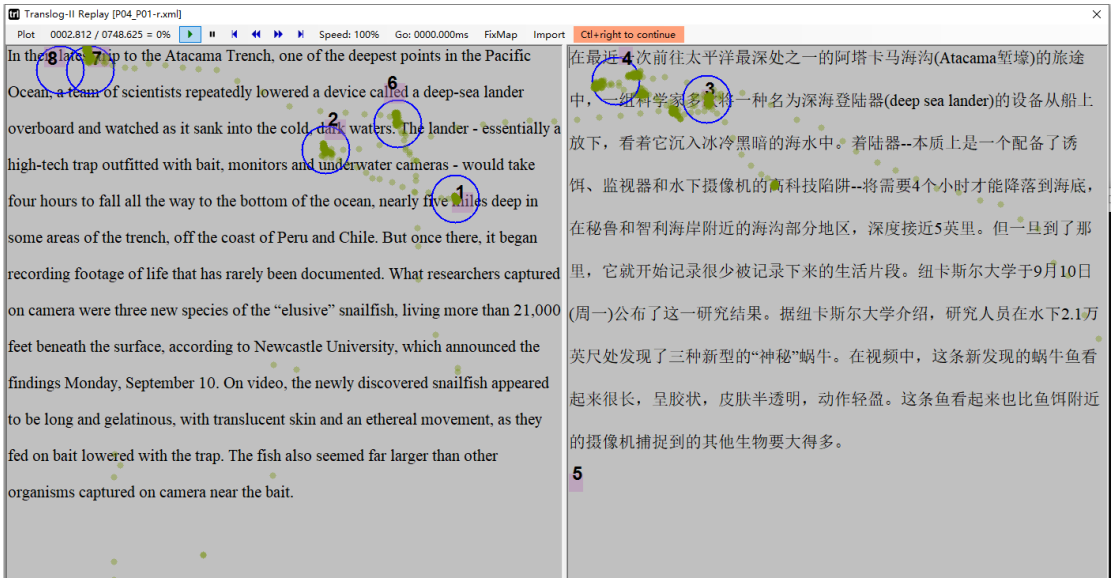
TPRDB 制作流程

一、前期操作

1. 前期工作（非常重要）

1.1 回放眼动数据，进行 offline gaze mapping（汉译英除外）



1.2 进行 manual fixation-to-word mapping

方法：使用 Fix Map 功能，利用 Tab 键切换光标位置。

注意：尽量不要修改横向的眼动数据。

标准：

Journal of Eye Movement Research
6(2):5, 1-11

Dynamic programming for re-mapping noisy fixations in translation tasks

Michael Carl
Copenhagen Business School

Eyetrackers which allow for free head movements are in many cases imprecise to the extent that reading patterns become heavily distorted. The poor usability and interpretability of these gaze patterns is corroborated by a "naïve" fixation-to-symbol mapping, which often wrongly maps the possibly drifted center of the observed fixation onto the symbol directly below it. In this paper I extend this naïve fixation-to-symbol mapping by introducing background knowledge about the translation task. In a first step, the sequence of fixation-to-symbol mappings is extended into a lattice of several possible fixated symbols, including those on the line above and below the naïve fixation mapping. In a second step a dynamic programming algorithm applies a number of heuristics to find the best path through the lattice, based on the probable distance in characters, in words and in pixels between successive fixations and the symbol locations, so as to smooth the gazing path according to the background gazing model. A qualitative and quantitative evaluation shows that the algorithm increases the accuracy of the re-mapped symbol sequence

Keywords: Fixation-to-symbol mapping, drift in gaze data, drift-correction algorithm

the first or second line - intuitively it seems more plausible that a translator reads source words which he or she is currently translating (as in Figure 2) instead of those words one line below (as in Figure 1). These observations lead us to the following criteria for a fixation-to-symbol re-mapping algorithm:

- successive fixations are more likely on neighbouring words than in the lines above or below
- translators are likely to read passages of source text words which they are currently translating
- the distance between the fixation center and the fixated characters should be minimal

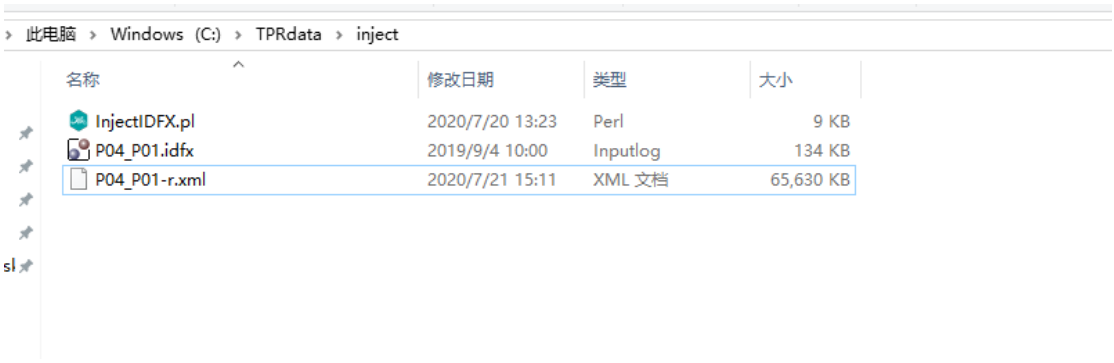
2. 增加语言对、任务标签

<Languages source="en" target="zh" task="post-editing" />
<Languages source="en" target="zh" task="translating" />

注：此步骤非常重要，因为会影响后续的眼动数据对齐效果和最后的 TPRDB 数据表的生成。

二、将 Inputlog 数据与 Translog 数据合并（如使用 Inputlog 收集被试网络搜索数据）

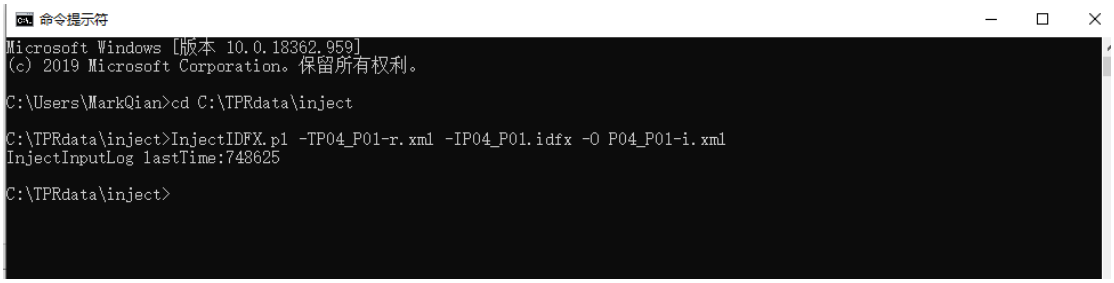
将.idfx、xml 和 InjectIDFX.pl 文件统一放入 Windows(C:)文件夹中。



然后执行以下代码：

C:\Users\MarkQian>cd C:\TPRdata\inject

InjectIDFX.pl -TP04_P01-r.xml -IP04_P01.idfx -O P04_P01-i.xml



三、制作 TPRDB

3.0 登录 TPRDB

https://critt.as.kent.edu/cgi-bin/yawat/yawat.cgi

用户名：JIAJUN

密码：jiajun

3.1 下载对齐文件(alignment file)

.atag; .src; .tgt

3.2 修正句段切分(segmentation)

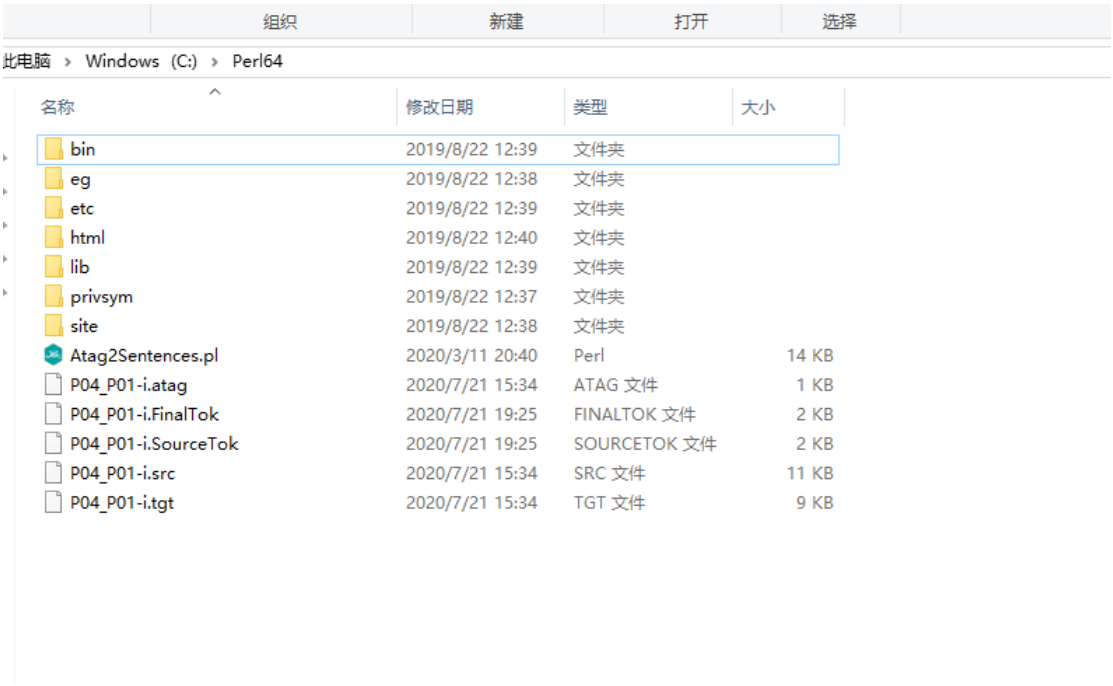
须安装 Perl 64，并将 Atag2Sentences.pl 和.atag; .src; .tgt 文件放在 C:\Perl64 目录下，然后执行以下代码：

C: \Users\MarkQian>cd C:\perl64

C:\Perl64>

C:\Perl64>Atag2Sentences.pl -A P04_P01-i

C:\Perl64>



用 Notepad++调整 segmentation，然后执行以下命令：

C:\Perl64>Atag2Sentences.pl -A P04_P01-i -o P04_P01-i

将修改过后的对齐文件重新压缩成 zip 包，上传至 TPRDB。

3.3 词对词对齐（如考察 Word Translation Entropy 或绘制翻译进程图）

Guidelines for Chinese-English Word Alignment

Version 4.0 – April 16, 2009

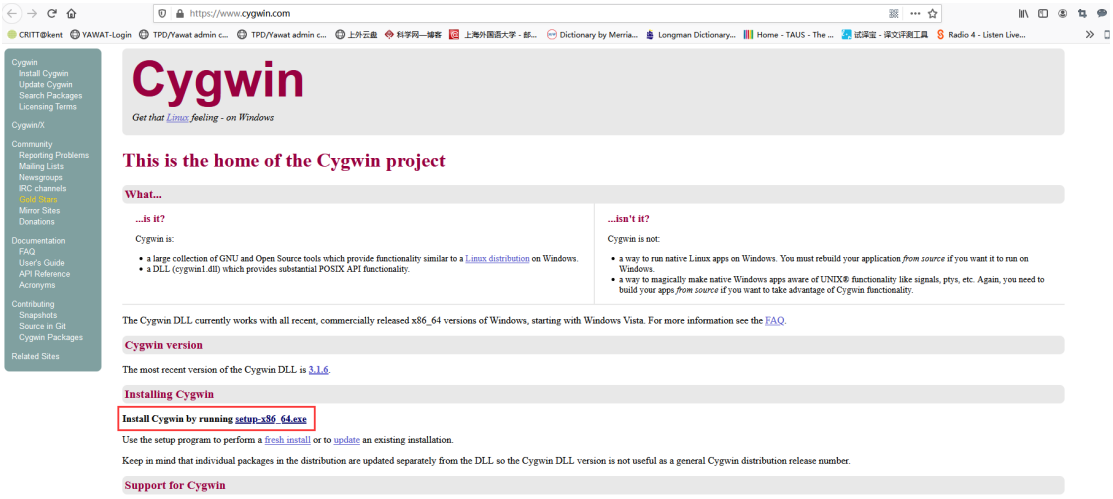
Linguistic Data Consortium

Created by: Xuansong Li xuansong@ldc.upenn.edu
With contribution from: Niyu Ge niyuge@us.ibm.com
Stephanie Strassel strassel@ldc.upenn.edu

www.ldc.upenn.edu/Project/GALE

四、绘制翻译进程图(Translation Progression Graph)

注：在 Cygwin(https://www.cygwin.com/)中操作




```
MarkQian@DESKTOP-VQ1D43A ~
$ cd /cygdrive/c/Users/MarkQian/Desktop/Jiajun19

MarkQian@DESKTOP-VQ1D43A /cygdrive/c/Users/MarkQian/Desktop/Jiajun19
$ for i in JIAJUN-tables/* ; do f=$(basename $i); echo "$i $f"; iconv.exe -f UTF-8 -t GB18030 $i >
```

Jiajun19_GB8030/$f; done

```
MarkQian@DESKTOP-VQ1D43A /cygdrive/c/Users/MarkQian/Desktop/Jiajun19
$ cd /cygdrive/c/Users/MarkQian/Desktop/Jiajun19

MarkQian@DESKTOP-VQ1D43A /cygdrive/c/Users/MarkQian/Desktop/Jiajun19
$ for i in JIAJUN-tables/* ; do f=$(basename $i); echo "$i $f"; iconv.exe -f UTF
-8 -t GB18030 $i > Jiajun19_GB8030/$f; done
JIAJUN-tables/JIAJUN.prot JIAJUN.prot
JIAJUN-tables/JIAJUN.ss JIAJUN.ss
JIAJUN-tables/P04_P01-i.ag P04_P01-i.ag
JIAJUN-tables/P04_P01-i.au P04_P01-i.au
JIAJUN-tables/P04_P01-i.ex P04_P01-i.ex
JIAJUN-tables/P04_P01-i.fd P04_P01-i.fd
JIAJUN-tables/P04_P01-i.fu P04_P01-i.fu
JIAJUN-tables/P04_P01-i.kd P04_P01-i.kd
JIAJUN-tables/P04_P01-i.pu P04_P01-i.pu
JIAJUN-tables/P04_P01-i.sg P04_P01-i.sg
JIAJUN-tables/P04_P01-i.st P04_P01-i.st
JIAJUN-tables/P04_P01-i.tt P04_P01-i.tt
JIAJUN-tables/P04_P01-r.ag P04_P01-r.ag
JIAJUN-tables/P04_P01-r.au P04_P01-r.au
JIAJUN-tables/P04_P01-r.ex P04_P01-r.ex
JIAJUN-tables/P04_P01-r.fd P04_P01-r.fd
JIAJUN-tables/P04_P01-r.fu P04_P01-r.fu
JIAJUN-tables/P04_P01-r.kd P04_P01-r.kd
JIAJUN-tables/P04_P01-r.pu P04_P01-r.pu
JIAJUN-tables/P04_P01-r.sg P04_P01-r.sg
JIAJUN-tables/P04_P01-r.st P04_P01-r.st
JIAJUN-tables/P04_P01-r.tt P04_P01-r.tt
JIAJUN-tables/progGra.R progGra.R

MarkQian@DESKTOP-VQ1D43A /cygdrive/c/Users/MarkQian/Desktop/Jiajun19
$ ^C
```

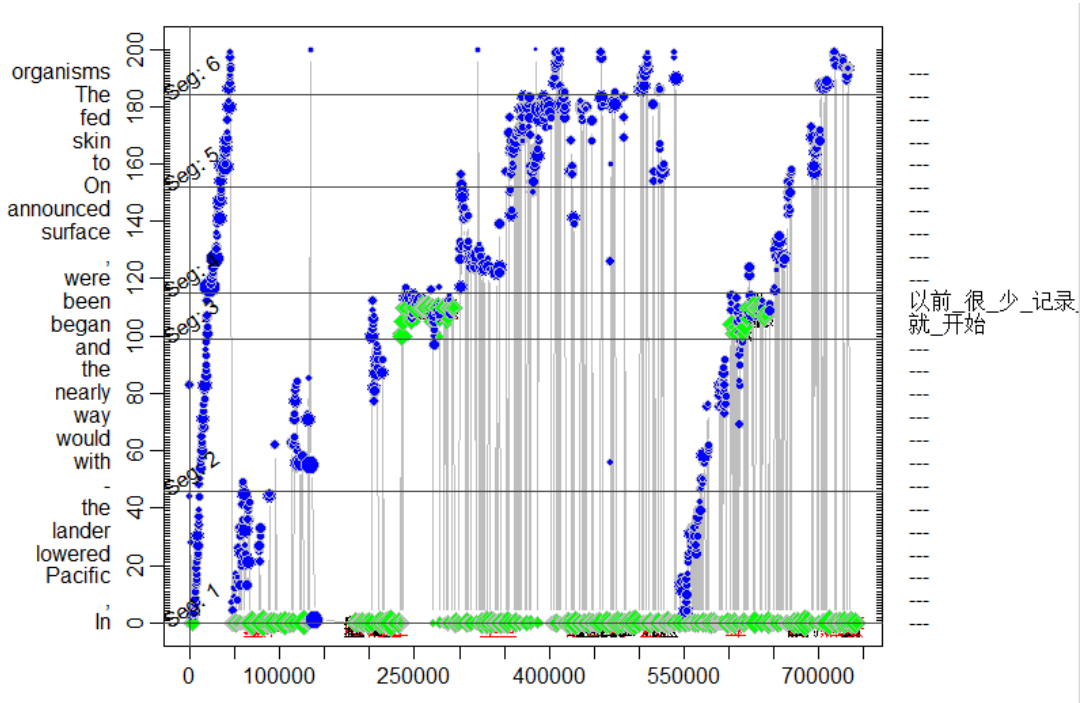
###2. Draw the translation progression graph

setwd("C:/Users/MarkQian/Desktop/Jiajun19/Jiajun19_GB8030")

source("C:/Users/MarkQian/Desktop/Jiajun19/Jiajun19_GB8030/progGra.R")

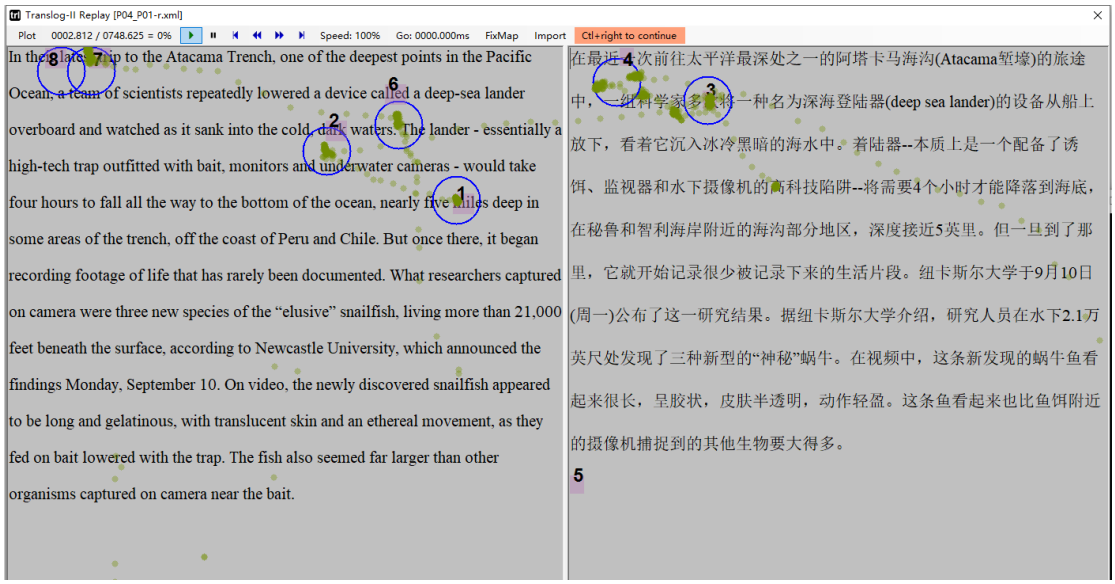
ReadData("C:/Users/MarkQian/Desktop/Jiajun19/Jiajun19_GB8030/P04_P01-i")

ProgGraph()



五、将数据表导入 R Studio

六、剔除噪音数据（transient fixation）



A heuristic-based approach for systematic error correction of gaze data for reading

Abhijit Mishra Michael Carl Pushpak Bhattacharya
(1) Indian Institute Of Technology, Bombay
(2)CRITT, Copenhagen Business School
(3) Indian Institute Of Technology, Bombay

abhijitmishra@cse.iitb.ac.in, mc.isv@cbs.dk, pb@cse.iitb.ac.in

ABSTRACT

In eye-tracking research, temporally constant deviations between users' intended gaze location and location captured by eye-samplers are referred to as systematic error. Systematic errors are frequent and add a lot of noise to the data. It also takes a lot of time and effort to manually correct such disparities. In this paper, we propose and validate a heuristic-based technique to reduce such errors associated with gaze fixations by shifting them to their true locations. This technique is exclusively applicable for reading tasks where the visual objects (characters) are placed on a grid in a sequential manner; which is often the case in psycholinguistic studies.

KEYWORDS: EYE-TRACKING, FIXATION CORRECTION, GAZE DATA MANIPULATION, SYSTEMATIC ERROR

2.2 Discarding transient fixations

Transient Fixations (TFs) are very short duration fixations which occur in between two fixations falling nearer to each other (on the same line or just a line apart) and located far away from each of them. In other words, upon joining three fixations if we observe a spike and the tip of the spike is a short duration fixation, it is said to be transient. Figure 1 illustrates one TF.

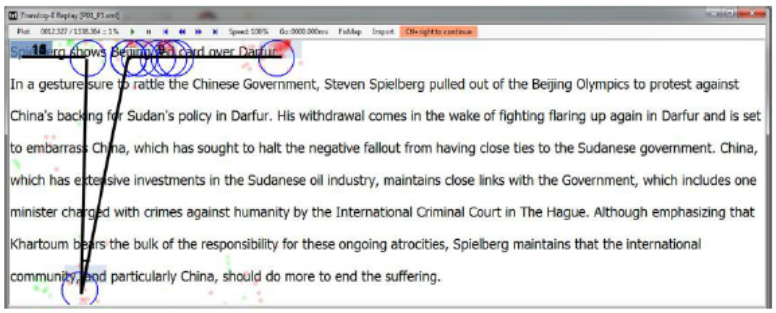


FIGURE 2 – Transient Fixations

Figure 2 shows one transient fixations. Upon joining 3 consecutive fixations involving one TF, we observe a spike.

P04_P01-i.fd - LibreOffice Calc

文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 工作表(S) 数据(D) 工具(T) 窗口(W) 帮助(H)

Liberation Sans 10

A6:AMJ6 4

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Id	Study	Session	SL	TL	Task	Text	Part	Time	Dur	Win	Cursor	X	Y	ParalK	Edit	EDid	Se	STid	TTid	
2	0	JIAJUN	P04_P01-i	en	zh	P	1	P04	77	348	1	406	569	220	0	---	---	2	83	0	
3	1	JIAJUN	P04_P01-i	en	zh	P	1	P04	421	172	1	215	405	158	0	---	---	1	44	0	
4	2	JIAJUN	P04_P01-i	en	zh	P	1	P04	624	324	2	44	887	94	0	---	---	1	0	22	
5	3	JIAJUN	P04_P01-i	en	zh	P	1	P04	952	418	2	3	772	71	0	---	---	1	0	3	
6	4	JIAJUN	P04_P01-i	en	zh	P	1	P04	1561	156	2	327	1982	1401	0	---	---	7	0	193	
7	5	JIAJUN	P04_P01-i	en	zh	P	1	P04	1921	180	1	143	496	125	0	---	---	1	28	0	
8	6	JIAJUN	P04_P01-i	en	zh	P	1	P04	2124	278	1	14	107	36	0	---	---	1	3	0	
9	7	JIAJUN	P04_P01-i	en	zh	P	1	P04	2405	436	1	6	69	6	0	---	---	1	2	0	
10	8	JIAJUN	P04_P01-i	en	zh	P	1	P04	2827	130	1	16	126	9	0	---	---	1	4	0	
11	9	JIAJUN	P04_P01-i	en	zh	P	1	P04	3014	192	2	46	914	90	0	---	---	1	0	24	
12	10	JIAJUN	P04_P01-i	en	zh	P	1	P04	3202	502	2	3	784	66	0	---	---	1	0	3	
13	11	JIAJUN	P04_P01-i	en	zh	P	1	P04	3749	590	1	15	120	9	0	---	---	1	3	0	
14	12	JIAJUN	P04_P01-i	en	zh	P	1	P04	4327	84	1	3	34	6	0	---	---	1	2	0	
15	13	JIAJUN	P04_P01-i	en	zh	P	1	P04	4421	236	1	11	89	13	0	---	---	1	3	0	
16	14	JIAJUN	P04_P01-i	en	zh	P	1	P04	5046	190	1	33	266	20	0	---	---	1	7	0	
17	15	JIAJUN	P04_P01-i	en	zh	P	1	P04	5233	340	1	39	311	37	0	---	---	1	8	0	
18	16	JIAJUN	P04_P01-i	en	zh	P	1	P04	5561	464	1	31	239	26	0	---	---	1	7	0	
19	17	JIAJUN	P04_P01-i	en	zh	P	1	P04	6030	320	1	38	304	24	0	---	---	1	8	0	
20	18	JIAJUN	P04_P01-i	en	zh	P	1	P04	6358	230	1	48	381	29	0	---	---	1	11	0	
21	19	JIAJUN	P04_P01-i	en	zh	P	1	P04	6577	288	1	56	442	29	0	---	---	1	13	0	
22	20	JIAJUN	P04_P01-i	en	zh	P	1	P04	6874	192	1	64	515	36	0	---	---	1	14	0	

For video demonstration, please contact Jiajun Qian <qianjiajun@shisu.edu.cn> for more information.

Jiajun Qian
Ph.D in Translation Studies
Shanghai Maritime University, China
ORCID: https://orcid.org/0000-0001-9143-0206
Website: https://www.researchgate.net/profile/Jiajun-Qian-2