



Teaching translation based on the mini-corpus Pedagogy and tool-assisted translation analysis

基于微型语料库的翻译教学：教学法与技术辅助的翻译学分析

Jun MIAO 繆君

miaoj5@mail.sysu.edu.cn





CONTENT

1

Corpus and translation teaching

语料库与翻译教学

2

Textometrics and mini-corpus

篇章词量学与迷你语料库

3

Classroom organization and student corpora

课堂组织与学生语料

4

Analysis and research findings

分析与研究结果

01

CORPUS AND TRASLATION TEACHING

语料库与翻译教学

FIRST PART



Big problem of translation teaching

Different interpretations of the same source text

- 1-How to compare their exercises ?
- 2-How to show them where the translation problem are?

evaluate => possibility in the translations



Corpus, corpora and our education

From the 1980s up to now

Corpora:

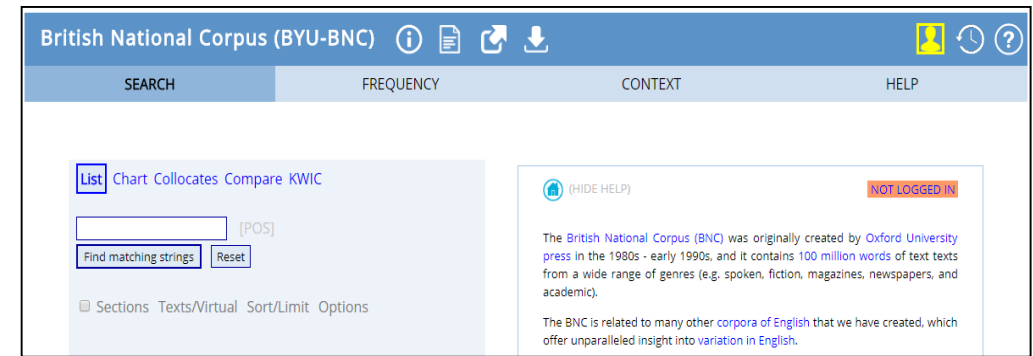
Increasingly vast, varied and accessible

University education:

Courses, discipline, department

Tools and research:

More varied and more efficient, various fields



100 million words



Teaching translation with tools



1. problems of corpus preparation (Bowker and Barlow 2008)
2. Insufficient application of corpus methods in pedagogical practice.

Pedagogical objectives



recherches

Large corpora
Methodology
Algorithms
Theory

Corpus size



practice

Small corpora
Theme
Market needs
Concrete exercises

Tool-based self-assessment and pedagogy

Corpus of students' homework/exercices

ICLE (*International Corpus of Learner English*), Sylviane Granger, University of Louvain, 1990

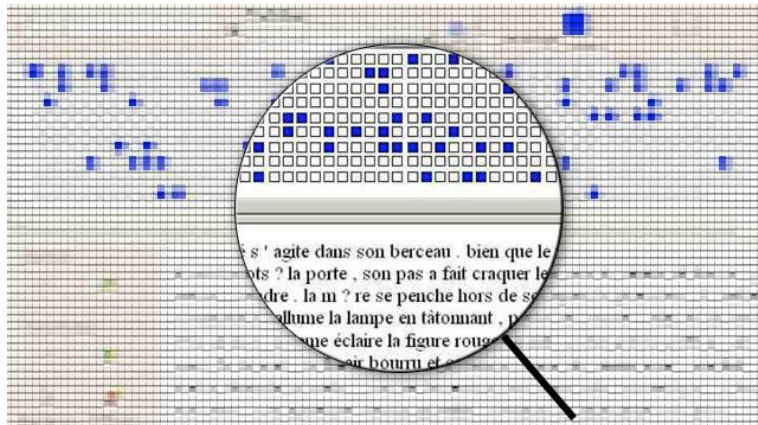
Student Translation Archive (Bowker & Bennison 2003)

ENTRAD (Florén 2006)

Russian Translation Learner Corpus (Sosnia 2006)

MeLLANGE (Castagnoli et al. 2009)

...



- Working in a group
- Self-assessment (Tercedor-Sánchez et al. 2005 ; Kiraly 2005 ; Valetopoulos 2012, etc)
- Translation quality
- To highlight differences between translations - students' problems

L'autoévaluation appuyée sur l'outillage textométrique dans
l'enseignement de la traduction
(在翻译教学中如何依托词量学工具对翻译进行自我评估)
Jun Miao & André Salem, 2016, N°61

02

TEXTOMETRICS AND MINI-CORPUS

篇章词量学与迷你语料库

SECOND PART



Discourse analysis

Michel Pêcheux (1938-1983) :

Automatic speech analysis (AAD69)

Paris, Dunod



In the 1960s

The question of ideology

The reading of ideological discourses

Quantitative objectivism

Philosophical structuralism

a) *Liste des énoncés*

	F	D ₁	N ₁	V	ADV	PP	D ₂	N ₂
ACU09411700000			S	IMPORTER	AUJ	A	O	NOUS
ACU09411800000			NOUS	SAVOIR	O	*	O	S
ACU09411900000			S	ETRE	O		O	IMAGINABLE
ACU0941200020L			PC	CHANGER	O	DE	O	ATTITUDE
ACU0941210020R			PC	CESSER	OTOT	O	O	O
ACU0941220000R			PC	CONSIDERER	O	*	LS	SOCIALISTE
ACU0941230000R			PC	CONSIDERER	O	COMM	SES	ADVERSAIRE
ACU0941240000R			ADVERSAIRE	E	O		O	PRINCIPAL
ACU0941250000R			PC	PREFERER	O	*	L	GOVERNEMENT
ACU0941260000R			GOVERNEMENT	E	O	DE	L	DROITE
ACU0941270000R			GOVERNEMENT	E	O	DE	L	CAPITAL
ACU0941280000R			CAPITAL	E	O		O	GRAND
ACU0941290000R			PC	PREFERER	O	A	L	VICTOIRE
ACU0941300000R			VICTOIRE	E	O	DE	LS	TRAVAILLEUR
ACU09413101000			RIEN	MONTRER	O	*	O	CELA

Statistics on language facts

Pierre Guiraud (years 50-60)

Jean-Paul Benzécri (1968, 1973, 1981...)

Maurice Tournier (1933-2013)

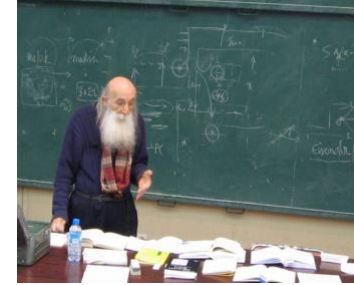
Charles Muller (1909-2015), statistique lexicale ou linguistique quantitative

Pierre Lafon (1977, 1984, 1992...) *Dépouillements et statistiques en lexicométrie*

André Salem (1982, 1984, 1987; 1991...2020)

« Pratique des segments répétés. Essai de statistique textuelle », INALF, 1987

« Méthodes de la statistique textuelle », thèse de doctorat d'Etat, Université Paris 3, 1993



Textometric tool -Lexico

History:

1970 : Laboratory of Mathematical Statistics, under Jean-Paul Benzécri

1980 : Labo of Saint-Cloud, Lexicometry et political textes, under Maurice Tournier

1990 : Syled, Paris 3, under André Salem

Lexico 5 :

version 5.8, since September/2017

<http://lexi-co.com/index.html>

utf-8, multilingual interface, more features



03

CLASSROOM ORGANIZATION AND STUDENT CORPORA

课堂组织与学生语料

THIRD PART



Course design and research method

In a translation training course



University Lyon 2 (2017-2018) : L3, 31 students, including 6 French students.

Sun Yat-sen University in Zhuhai (2019-2020) : M1,

77 students : 21 students of interpreter class, 56 students of translation class.

A total : 98 students

My translation

Automatic translations :

- DeepL : <https://www.deepl.com/translator>
- Google translation : <https://translate.google.fr/>
- Baidu: <https://fanyi.baidu.com/>
- Jinshan: <http://fy.iciba.com/>
- Youdao: <http://fanyi.youdao.com/>
- Iflytek: <https://www.iflyrec.com/html/translate/upload.html>

To encourage students to discover themselves the phenomena in the field of translation



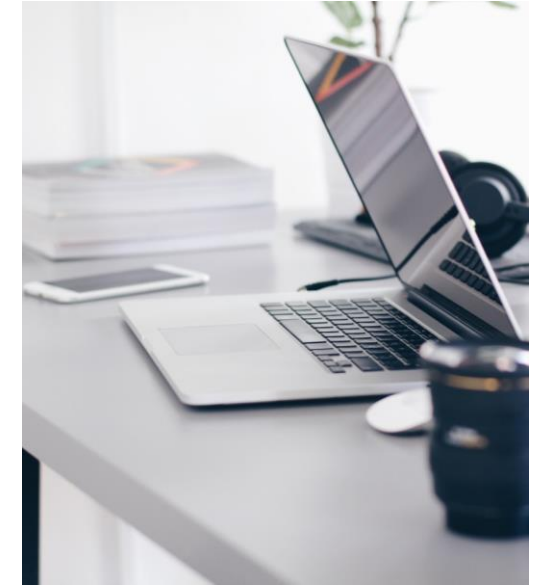
Small corpora, mini-corpus

We insist on the interest of the corpus :

- of reduced size ;
- carry out thematic and detailed research;
- calculation methods +/-: absolute frequency...

Aims :

- facilitate thematic analysis, methodology
- reducing subjectivity, highlighting facts
- the pragmatic method, from bottom to top
- **corpus-driven vs. corpus-Based**
- **empirical approach**



Two sentences of Jean Nouvel 让·努维尔



30 words :

I can assure you that I will make myself totally available for considering your needs and desires and find with you the best possible answers.

My architecture will be yours.



Processing of the corpora

1 Anonymization of student' names

2. Chinese segmentation 中文切分 (*NLPIR-ICTCLAS*, Zhang Huapin 张华平, <http://ictclas.nlpir.org/>)

3. Structuring the corpus :

```
#<version=01A02> 您放心，我将为您的需求与期待全情投入，  
#与您共寻最佳方案。  
#您所想即我所建。  
#<version=01B03> 请放心，我将考虑您的所有需求，让此建筑尽可能使您满意，  
#虽出自我手，但走入您心。  
#<version=01B04> 我向您保证，我会全盘考虑您的需求和愿望，并与您一道找出最佳可行方案。  
#我的建筑将属于您。  
#<version=01C05> 我会全心全意考虑您的需求与愿望，为您找到最优方案。  
#我的设计为您存在。  
#<version=01C06> 我向您保证，我会百分百考虑您的需求和愿望，为您提供最优方案。  
#我的设计为您而生。  
#<version=01D07> 我向你保证，我会充分考虑你的愿望和需求，并为你找到最佳答案。  
#它是我的设计，是你的建筑。  
#<version=01D08> 我保证，我将竭我所能，依你所求，如你所愿，同你一道找寻最佳答案。  
#这是我的建筑，也是你的。  
#<version=01E09> 我保证会根据你的所需和所想给出尽可能让你满意的答复。  
#你可以用我的建筑式样。  
#<version=01E10> 我保证会深思熟虑你的需求和愿望，我会和你一同寻找最佳方案。  
#我的建筑也是你的建筑。  
#<version=01F11> 我向您保证，我定会充分考虑您的需求 & 满足您的期望，和您一起找到最佳方案。  
#我的设计为您而生。
```

Mini-corpora JN

04

ANALYSIS AND RESEARCH FINDINGS

分析与研究结果

FOURTH PART



Four parts of our mini-corpus JN

N	Part	Tokens	Types	Hapax	Fmax	Type	Occurrences
1	autotrans	190	36	7	25	的	
2	lyon2	944	191	98	95	的	
3	myself	28	23	20	4	的	
4	sunyatsen	2147	290	141	226	您	
T	Corpus	3309	345	150	322	您	

N	Part	Tokens	Types	Hapax	Fmax	Type	Tokens
74	02ZAA53	26	21	18	4	我	
75	02ZAA54	35	27	24	5	的	
76	02ZBB55	31	23	20	4	您	
77	02ZBB56	30	22	16	4	您	
78	ABaidu	31	23	20	4	的	
79	ADeepI	32	24	20	4	的	
80	AGoogle	31	21	16	4	的	
81	Alfly	30	22	18	4	的	
82	AJinshan	33	23	18	4	的	
83	AYoudao	33	24	21	5	的	
84	LF0107	28	23	20	4	的	
85	LF0208	28	23	20	4	的	
86	LF0309	28	19	13	4	的	
87	LF0416	28	20	16	4	的	
88	LF0525	34	23	18	5	的	
89	LF0628	35	25	20	5	的	
90	LZ0101	30	24	20	4	的	
91	LZ0202	30	23	19	4	您	
92	LZ0303	34	26	21	4	的	
93	LZ0404	25	20	17	3	您	
94	LZ0505	31	26	23	3	您	
95	LZ0606	31	24	20	4	您	
96	LZ0710	35	26	21	5	您	
97	LZ0811	30	22	18	4	您	
98	LZ0912	29	21	17	4	您	
99	LZ1013	30	28	26	2	您	
100	LZ1114	26	21	18	3	您	
101	LZ1215	17	16	15	2	汝	
102	LZ1317	24	21	19	3	您	
103	LZ1418	33	26	23	4	您	
104	LZ1519	32	25	21	4	您	
105	LZ1620	30	21	17	4	您	
106	LZ1721	28	24	22	4	您	
107	LZ1822	33	27	23	3	您	
108	LZ1923	30	21	15	4	的	
109	LZ2024	37	27	22	4	您	
110	LZ2126	31	24	20	4	的	
111	LZ2227	31	27	24	3	你们	
112	LZ2329	33	23	17	4	您	
113	LZ2430	39	31	28	6	的	
114	LZ2531	34	28	25	3	的	
115	Miaojun	28	23	20	4	的	
T	Corpus	3309	345	150	322	您	

	N	Part	Tokens	Types	Hapax	Fmax	Type	Tokens
✓	28	02D07	25	18	14	4	您	
✓	29	02D08	38	30	27	4	的	
✓	30	02E09	20	16	14	4	您	
✓	31	02E10	20	19	18	2	的	
✓	32	02F11						
✓	33	02F12						
✓	34	02G13						
✓	35	02G14						
✓	36	02H15						
✓	37	02H16						
✓	38	02I17						
✓	39	02I18						
✓	40	02J19						
✓	41	02J20						
✓	42	02K21						
✓	43	02K22						
✓	44	02L23						
✓	45	02L24						
✓	46	02M25						
✓	47	02M26						
✓	48	02N27						
✓	49	02N28						
✓	50	02O29						
✓	51	02O30						
✓	52	02P31						
✓	53	02P32						
✓	54	02Q33						
✓	55	02Q34						
✓	56	02R35						
✓	57	02R36						
✓	58	02S38						
✓	59	02T39	20	20	10	3	您	
✓	60	02T40	29	21	17	4	您	
✓	61	02U41	40	30	23	4	你们	
✓	62	02U42	28	23	20	4	您	
✓	63	02V43	29	22	19	4	您	
✓	64	02V44	22	19	16	2	的	
✓	65	02W45	31	24	21	4	您	
✓	66	02W46	27	22	19	3	的	
✓	67	02X47	28	20	16	4	您	
✓	68	02X48	30	23	19	4	的	

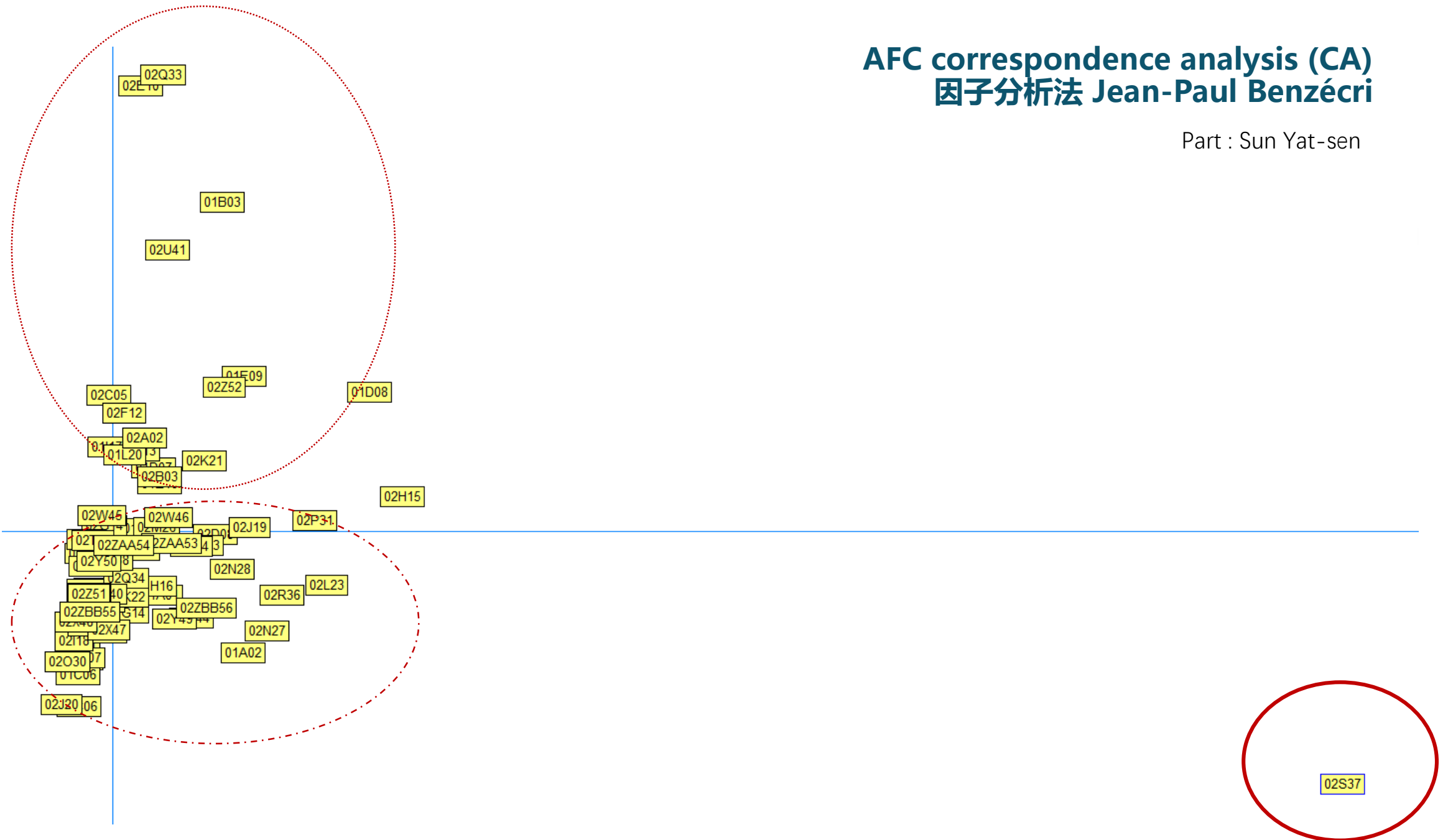
Context length: Tokens: 40 Continuous parts: oui Enregistrer

<version=02U41>我向你们保证，我将尽心尽力把你们的需求和愿望考虑到设计中，和你们一起找出最佳的待选方案。你们所指便是这座建筑设计的所向。¶

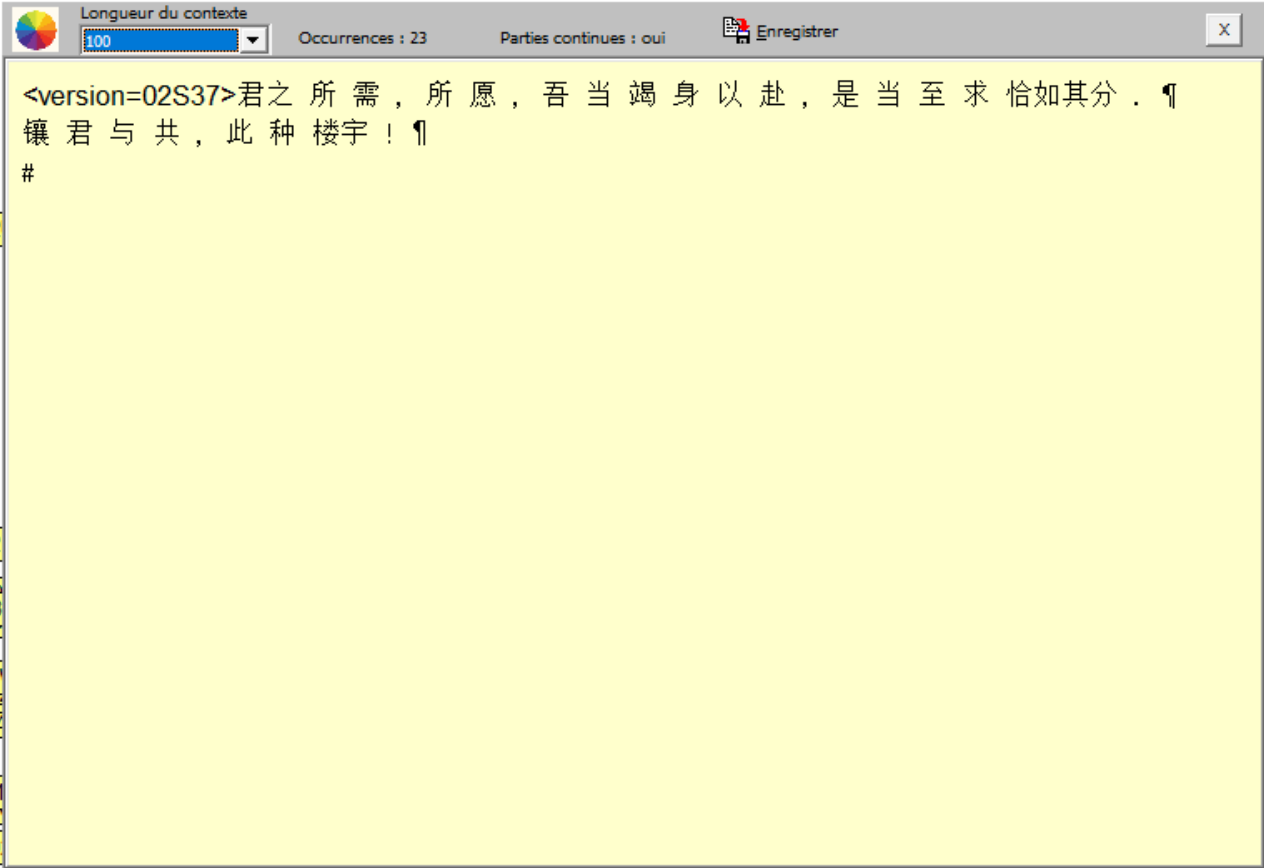
#

AFC correspondence analysis (CA) 因子分析法 Jean-Paul Benzécri

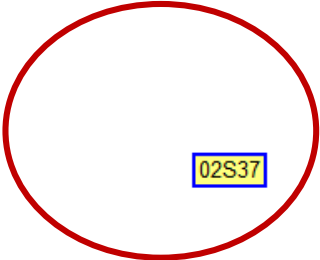
Part : Sun Yat-sen



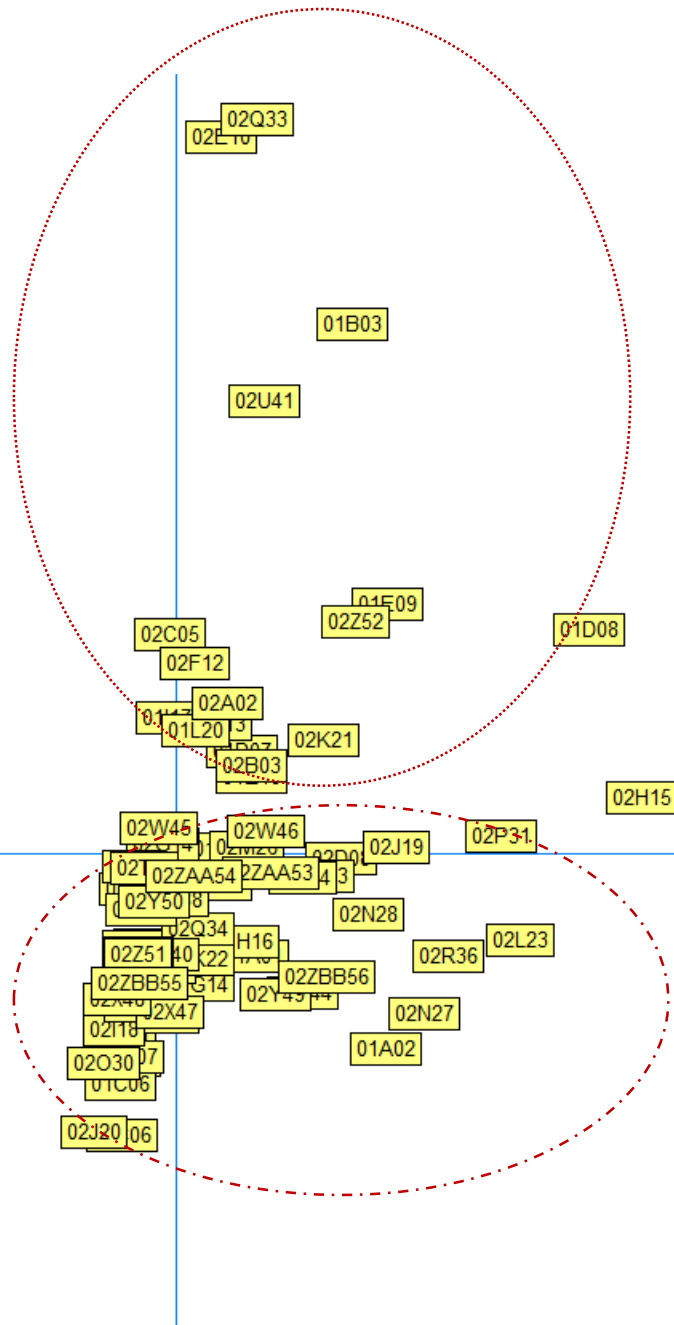
Return to the original text



The screenshot shows a window titled "Longueur du contexte" with a search bar containing "100". The window displays the following text: `<version=02S37>`君之所需，所愿，吾当竭身以赴，是当至求恰如其分。¶
镶君与共，此种楼宇！¶
#



CA and specificities calcul



Terme	F ...	f ...	S ...
你	45	32	12
你们	10	9	6
是	19	11	4
为	40	3	-3
您	226	19	-10

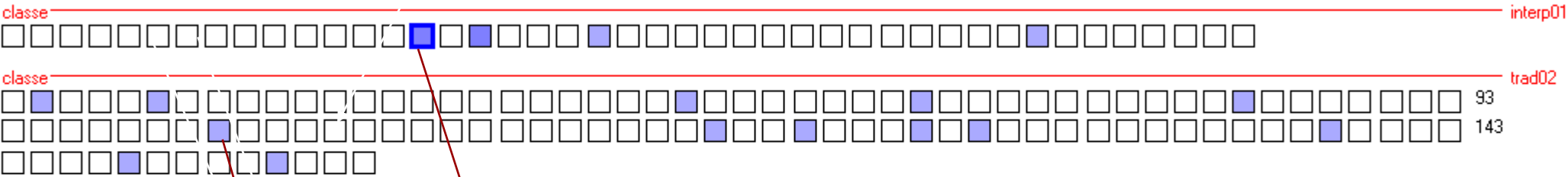
1-personal pronoun
2-copula 是
3-verb 为

Terme	F ...	f ...	S ...
您	226	207	11
为	40	37	3
是	19	7	-4
你们	10	1	-6
你	45	13	-11

Section map and display

The original: My architecture will be yours.

Sections 19 Partition classe [Seuillage] [Effacer] [Exporter]



Sections Occurrence Rapport

#它是我的设计,是你的建筑 .¶
It's my design, (it) is your architecture.

Sections Occurrence Rapport

#我设计的建筑将会是您的理想型 .¶
The building I design will be your ideal type .

Concordance of 是

< left Key... right >

-- Part : interp01, Number of contexts : 6

思熟虑你的需求和愿望,我会和你一同寻找最佳方案.¶我的建筑也,依你所求,如你所愿,同你一道找寻最佳答案.这是我的建筑,也可向你保证,我会充分考虑你的愿望和需求,并为你找到最佳答案.它会充分考虑你的愿望和需求,并为你找到最佳答案.它是我的设计,我将竭我所能,依你所求,如你所愿,同你一道找寻最佳答案.这提供最佳方案.我的建筑,您的选择.¶#我向您保证,满足您的需求

是
是
是
是
是

你的建筑.¶#我向您保证,我定会充分考虑您的需求&满足您的期望,和您的.¶#我保证会根据你的所需和所想给出尽可能让你满意的答复.¶我的设计,是你的建筑.¶#我保证,我将竭我所能,依你所求,如你所求.¶#我保证,我将竭我所能,依你所求,如你所愿,同你一起.¶我的建筑,也是你的.¶#我保证会根据你的所需和所想给出尽可能让我我的愿望,我将与您一起找到最好的答案.好的建筑,您值得拥有.¶#我

-- Part : trad02, Number of contexts : 13

全力满足您的需求和渴望,与您共同探寻最佳方案.我的建筑设计亦会全心意想你所需&想你所想,为您寻找最佳设计.我设计的建筑将会并行考虑,并和您一起寻找最好的可能的答案.最终的建筑成品将会我将完全尊重您的需求和喜好,并为您提供最佳方案.¶我的设计会会完全考虑你的需求和心愿,与你一同寻找最佳答案.我的建筑,你愿望考虑到设计中,和你们一起找出最佳的待选方案.你们所指便会全心全意关注您的需求和愿望,并为您提供最佳选择.我的建筑即风格.¶#竭我之力想你所想,倾我所能予你所求.¶我的建筑也将注意,考虑您的需求和愿望,并与您携手探索最佳方案.¶我的建筑将需要和愿望,并与您一同找寻最理想的建筑设计.¶我设计的建筑将答案.¶阿布扎比卢浮宫,终属于您.¶#满足您的需求,实现您的愿望,心愿,并且为您提供最优选择.我的设计,为您而生.¶#我能确保的最佳方案.我之所筑为你所属.¶#君之所需,所愿,吾当竭身以赴,

是
是
是
是
是
是
是
是
是
是
是
是

您的作品.¶#我能向你保证说我会充分考虑你的需求与意愿,并给你您的理想型.¶#我向您保证,我将充分考虑您的需求愿望,与您一起找到您心中的模样.¶#我向您保证我会用心地考虑您的需求和愿望,并为您提供最适合您的设计.¶#我向你们保证,我将尽心尽力把你们的需求和愿望奉主人.¶#我可以向您保证,我将全心全意地考虑您的需求和愿望,与您共这座建筑设计的所向.¶#我向您保证,一定会充分考虑您的需求和愿望您的建筑.¶#客户请放心,我会充分考虑您的需求和期望,与您共同探求您的作品.¶#我向您保证,我将全心全意考虑您的需求和渴望,与您一起您的心之所向.¶#我向你保证,我完全将你的需求和愿望纳入考虑,并您心中的理想建筑.¶#我保证,将全力以赴,满足您的需求和愿望,与您并我郑重的承诺.让我们共寻答案,共享建筑之美.¶#我可以向你保证,我,我将完全把您的需求和意愿置于首位,与您一起寻得最佳方案.最终呈当至求恰如其分.¶#君与共,此种楼宇!¶#我向您保证,我会充分考虑

Different translations for "will be" (是)

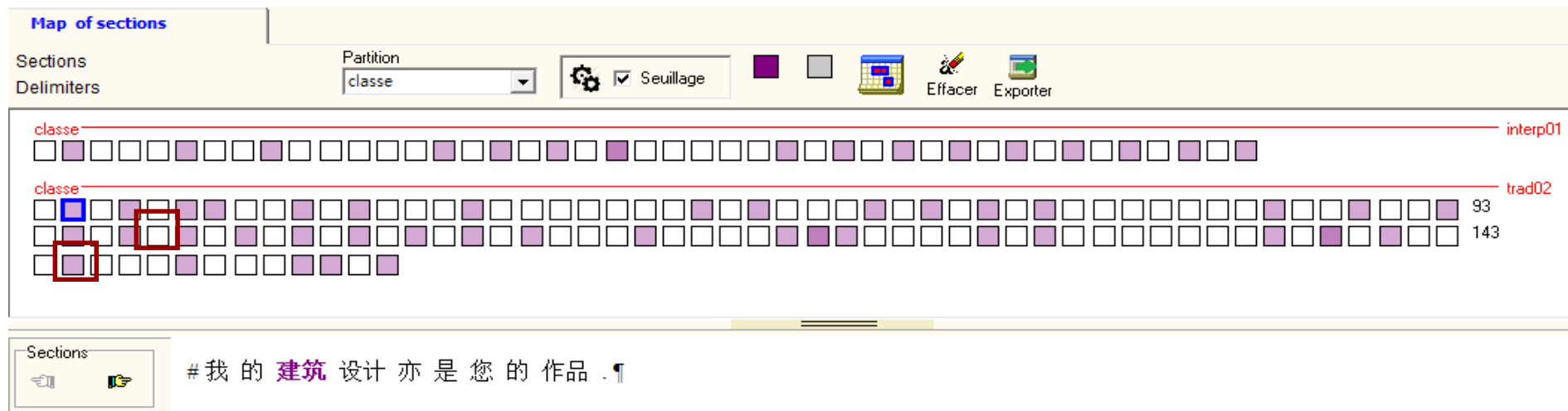
- Three major translations:

1- Express the meaning of "at the same time", "also" 表达“同时”、“同样”含义: 也是, 亦是, 即是, 便是

2- Express implicitly the meaning of "reality" 隐晦表达“现实”意义: 是, 这是

3- close to the original text when expressing "will be" 贴近原文, 表达“将来”时: 将会是, 将是

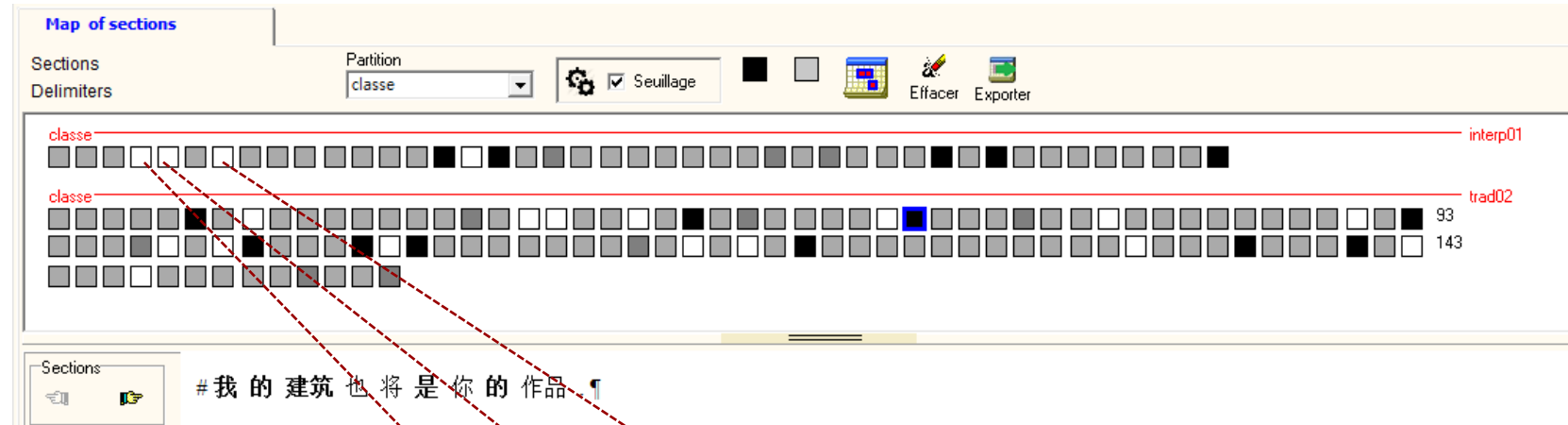
How to translate “architecture”?



建筑	58	楼宇	1
设计	10	构建	1
作品	4	->量身打造	1
成果	1	其他意译	3
所筑	1		

Repeated segments+sepecitifiy and locatisation

Term	F	f	S
建筑	58	58	37
的建筑	45	45	29
我的建筑	24	24	16
我的	40	25	7
建筑将	10	10	7
是	19	14	6
的建筑将	9	9	6
的	180	69	5
是您	6	6	5
建筑设计	7	7	5
的建筑设计	5	5	4
设计的	5	5	4
理想	3	3	3
您的建筑	3	3	3
将是	3	3	3
属于您	8	5	3
最终	3	3	3
属于	11	7	3
也将	3	3	3
心中	3	3	3
我设计的建筑将	4	4	3
我的建筑将	4	4	3
是您的	4	4	3
我的建筑设计	4	4	3
将	55	22	3
也	7	5	3
是你的	4	4	3
设计	30	14	3
由	5	4	3
建筑之美	2	2	2

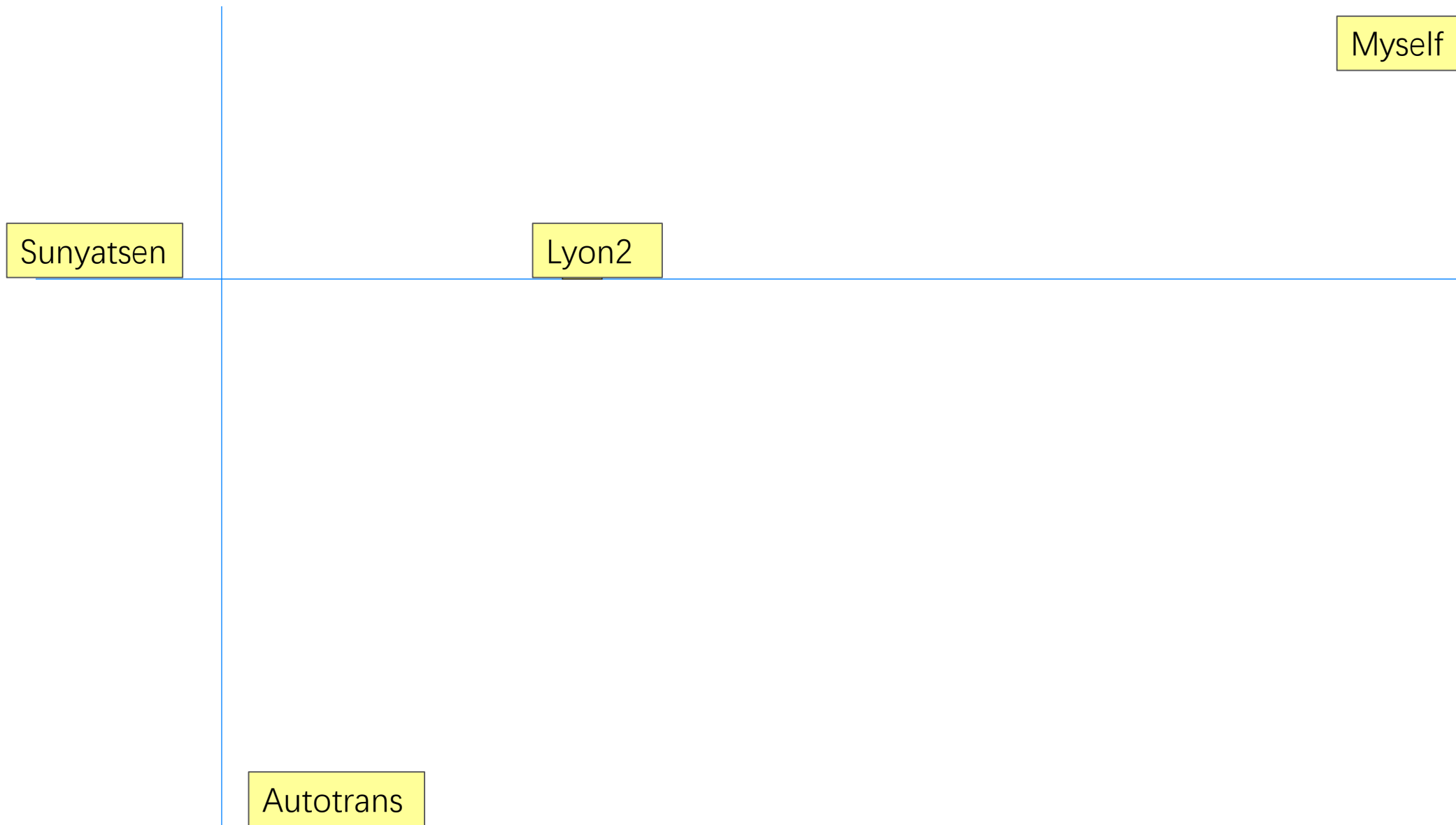


My architecture will be yours.

- # 与您共寻最佳方案.
- # 您所想即我所建.
- # 虽出自我手,但走入您心.
- # 总有一款适合你.
- # 你值得拥有
- # 我之所筑为你所属.
- # 镶君与共,此种楼宇!
- # 精彩,量身打造.
- # 我倾尽所能,将成果献给您.
- # 阿布扎比卢浮宫,终属于您.

Correspondence analysis (CA)

4 parts of corpus JN



AC+specificities

Sunyatsen

Lyon2

Myself

Specificities positive negative

Term	F ...	f ...	S ...
意愿	18	13	5
同	14	11	5
解决	14	11	5
投入	11	8	4
并	50	21	3
你	75	12	-3

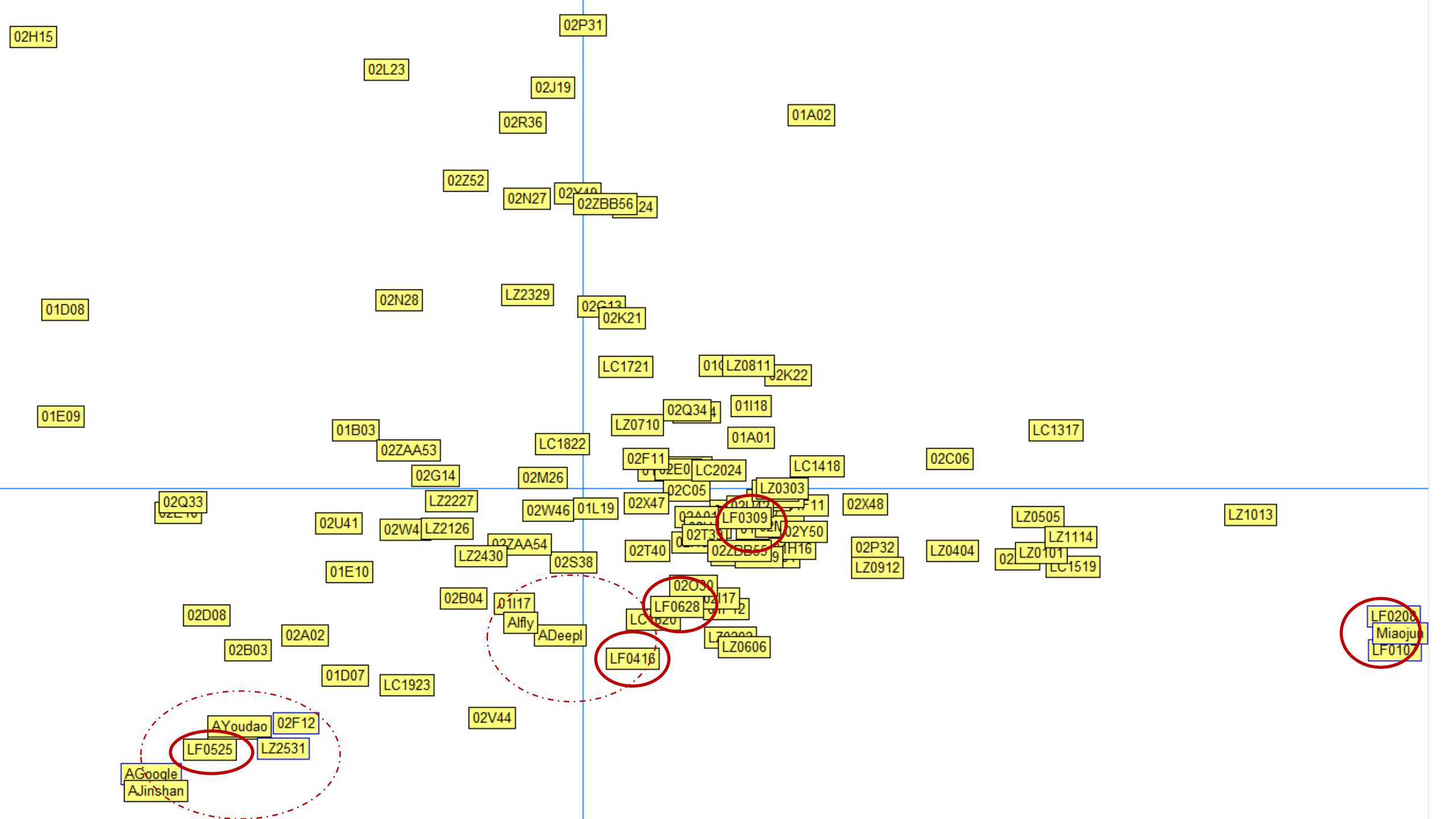
Specificities positive negative

Term	F ...	f ...	S ...
您	322	226	3
所	56	44	3
寻找	13	12	3
属于	11	11	3
放心	10	10	3
和	84	46	-2
的	304	180	-3
找到	41	19	-3
可以	29	13	-3
并	50	22	-4
投入	11	2	-4
意愿	18	4	-5
同	14	2	-5
解决	14	2	-5

Autotrans

Specificities positive negative

Term	F	f	S
你	75	18	8
可以	29	8	5
好	23	6	4
是	29	6	3
答案	27	6	3
的	304	25	2
最	43	6	2
您	322	6	-5



02H15

02P31

02L23

02J19

02R36

01A02

02Z52

02N27

02Y40

02ZBB56

24

01D08

02N28

LZ2329

02G13

02K21

LC1721

01LZ0811

02K22

01E09

01B03

02ZAA53

LC1822

LZ0710

02Q34

01I18

01A01

LC1317

02C06

02U41

02G14

02M26

02F11

02E0

LC2024

LC1418

02Q33

02L10

LZ2227

02W46

01L19

02X47

02C05

LZ0303

02X48

LZ1013

02U41

02W4

LZ2126

02ZAA54

02S38

02T40

02T3

02Y50

02P32

LZ0404

02LZ0101

LC1519

01E10

02B04

01I17

02O30

LF0628

02I17

LC020

LF0418

LZ0606

02D08

02B03

02A02

01D07

LC1923

02V44

AYoudao

02F12

LF0525

LZ2531

AGoogle

AJinshan

LF0200

Miaoju

LF0100

Translations from automatic translation vs. my translation

#<partie=autotrans>

#<version=AGoogle>我可以向你保证,我将使自己完全可以考虑你的需求和愿望,并找到你最好的答案。

#我的架构将是你的。

<version=ADeepI>

我可以向您保证,我将会全力以赴地考虑您的需求和愿望,并与您一起找到最好的答案。

我的建筑将是你的。

<version=ABaidu>

我可以向你保证,我会充分考虑你的需要和愿望,并与你一起找到最好的答案。

我的建筑将是你的。

<version=AJinshan>

我可以向你保证,我将使自己完全可以考虑你的需要和愿望,并与你找到最好的答案。

我的建筑将是你的。

<version=AYoudao>

我可以向你保证,我会充分考虑你的需要和愿望,并与你一起找到最好的可能的答案。

我的建筑将是你的。

<version=AIfly>

我可以向您保证,我将随时考虑您的需求和愿望,并与您一起找到最好的答案。

我的建筑将是你的。

#<partie=myself>

#<version=Miaojun>我将全力投入,研究分析贵方的需求和意愿,并同您一起,找到最佳的解决方案。

#我的设计必将成就您的建筑。

Use of idioms

DeepL

全力以赴	1
------	---

Lyon2 part

0 French students

全心全意	2	舍身处地	2
竭尽全力	2	全心全力	2
全力以赴	3		

Sun Yat-sen part

全心全意	10	深思熟虑	1
竭尽全力	3	设身处地	1
尽心尽力	2	有朝一日	1
恰如其分	1	魂牵梦萦	1
宵衣旰食	1		

<version=01E10>我保证会深思熟虑你的需求和愿望,我会和你一同寻找最佳方案. ¶

我的建筑也是你的建筑. ¶

05

CONCLUSION

结论

LAST PART



PEDAGOGY

self-learning
observation
comparison
work in groups

TEXTOMETRICAL TOOL

huge corpora or mini-corpus
quantitative and qualitative analysis
terminology
Support for automatic translation

TRANSLATOLOGICAL
ANALYSIS

systematic method of analysis
divergence => why
translation quality
translation assessment criteria
foreign language learners:
synonym, idiom



Thank you for your attention

Merci de votre attention

感谢大家的关注

Questions 问题 ?

Jun MIAO 繆君

miaoj5@mail.sysu.edu.cn

