



上海交通大学  
SHANGHAI JIAO TONG UNIVERSITY



上海交通大学外国语学院  
School of Foreign Languages

# 第五届中国语料库语言学 大会：思想与技术

## 会议日程

中国英汉语比较研究会语料库语言学专业委员会

上海交通大学外国语学院

2022. 11. 26-27



# 目 录

- 一、 会议简介
- 二、 日程安排
- 三、 主旨发言
- 四、 宣读论文



## 一、会议简介

语料库语言学研究，栉风沐雨六十载，春华秋实一甲子。新时代，语言数据科学的大发展给语料库研究带来新的机遇与挑战。语料库语言学研究，不仅要有不断革新的技术，还要有引领发展的新思想和新理念；更重要的是，诚如辛克莱（John Sinclair）所言，要对语言事实保持思想开放。在此背景下，“**第五届中国语料库语言学大会：思想与技术**”将于2022年11月26-27日在上海举办。届时诚邀广大语料库研究者参加，一起交流研究成果，碰撞学术思想火花，共同助力语料库语言学的新发展。我们将针对如下主题展开深入探讨：语料库语言学思想探究、语料库与文本意义分析、语料库与数据科学应用研究、新型语料库技术开发与应用、多语汉外平行语料库建设与应用、语料库语言学视角下的数字人文研究、语料库驱动的话语分析、语料库与教学研究、语料库与翻译研究、语料库与文学研究等。大会由中国英汉语比较研究会语料库语言学专业委员会主办，上海交通大学外国语学院承办。



## 二、 日程安排

2022 年 11 月 26 日上午		
腾讯会议：941258888      密码：221126 Bilibili 直播间： <a href="https://live.bilibili.com/26167061">https://live.bilibili.com/26167061</a>		
时间	内容	主持人
开幕式		
9:00-9:30	1. 上海交通大学 <b>杨惠中教授</b> 致辞 2. 中国英汉语比较研究会语料库语言学专业委员会会长 <b>梁茂成教授</b> 致辞 3. 上海交通大学外国语学院院长 <b>常辉教授</b> 致辞 4. 上海交通大学“语料库与数据科学研究中心”成立揭牌仪式 5. 线上合影	<b>甄凤超 教授</b> 上海交通大学
主旨发言		
9:30-10:30	<b>卫乃兴教授</b> 北京航空航天大学 题目：Corpus Approaches to Discourse Studies: methodologies, Challenges and Prospects	<b>许家金 教授</b> 北京外国语大学
10:30-11:30	<b>李文中教授</b> 浙江工商大学 题目：Textual Data Science for Corpus Linguistic Exploration of Textual Objects and Their Paraphrases	<b>濮建忠 教授</b> 浙江工商大学
2022 年 11 月 26 日下午		
分论坛讨论		
时间	内容	主持人
13:30-17:10	<b>分论坛 1：基于语料库的学术语言研究</b> 腾讯会议：927430820 密码：174722	<b>姜峰 教授</b> 吉林大学 <b>林玲 副教授</b> 上海交通大学
	<b>分论坛 2：基于语料库的话语分析</b> 腾讯会议：992709286 密码：435435	<b>钱毓芳 教授</b> 浙江工商大学 <b>郇昌鹏 副教授</b> 上海交通大学
	<b>分论坛 3：基于语料库的语言对比、学习者语言研究</b>	<b>邵斌 教授</b> 浙江大学



	腾讯会议：631675611 密码：637053	吴淑琼 教授 四川外国语大学
2022 年 11 月 27 日上午		
分论坛讨论		
时间	内容	主持人
8:30-11:50	<b>分论坛 4：基于语料库的翻译研究</b> 腾讯会议：325325560 密码：908388	<b>黄立波 教授</b> 西安外国语大学 <b>庞双子 长聘副教授</b> 上海交通大学
	<b>分论坛 5：基于语料库的汉语研究</b> 腾讯会议：580624065 密码：803387	<b>郭鸿杰 教授</b> 上海财经大学 <b>郭曙纶 副教授</b> 上海交通大学
	<b>分论坛 6：语料库文体学研究</b> 腾讯会议：471334614 密码：456653	<b>孙海燕 教授</b> 河南师范大学 <b>李鑫 副教授</b> 上海交通大学
	<b>分论坛 7：语料库技术研究</b> 腾讯会议：466890931 密码：360779	<b>陆军 教授</b> 扬州大学 <b>管新潮 副教授</b> 上海交通大学
2022 年 11 月 27 日下午		
腾讯会议：919455296      密码：221126 Bilibili 直播间： <a href="https://live.bilibili.com/26167061">https://live.bilibili.com/26167061</a>		
主旨发言		
13:30-14:30	<b>邓耀臣教授</b> 大连外国语大学 题目：数字人文研究述评：热点领域与研究趋势	<b>雷蕾 教授</b> 上海外国语大学
14:30-15:30	<b>Michaela Mahlberg 教授</b> 伯明翰大学 题目：(Corpus) Linguistics in the Digital Age	<b>甄凤超 教授</b> 上海交通大学
15:30-16:30	<b>Wolfgang Teubert 教授</b> 伯明翰大学 题目：Why Discourse is Not an Object of Science	<b>娄宝翠 教授</b> 河南师范大学
16:30-16:40	闭幕式	<b>甄凤超</b>



**分论坛 1: 基于语料库的学术语言研究** (腾讯会议: 927430820; 密码: 174722 )

**主持人:** 姜峰教授 吉林大学; 林玲副教授 上海交通大学

序号	时间	报告人	题目	主持人
1	1:30-1:50	刘传江 (广东金融学院)	学术英语词表验证的共选路径——以COMMON 为例	姜峰教授
2	1:50-2:10	耿文殊 (北京航空航天大学)	A quantitative approach for polysemous meaning identification in academic paper	
3	2:10-2:30	韩超 (兰州大学)	The use of Transition Markers in Chinese and English Student Academic Writing	
4	2:30-2:50	张畅 (香港教育大学)	A Corpus-based study of hedges in the academic oral presentation	
5	2:50-3:10	崔云凯 (烟台南山学院)	外国语言文学类期刊英文摘要立场标记语研究	
6	3:10-3:30	郭思文 (北京航空航天大学)	基于信息熵的科技语篇和科普语篇体裁特征研究	林玲副教授
7	3:30-3:50	邢富坤 (浙江外国语学院) 仲可欣 (青岛大学)	英文学术论文摘要的小句复合体特征研究	
8	3:50-4:10	孙淑怡姜峰 (吉林大学)	A corpus-based study of negation in PhD theses' limitations sections across disciplines	
9	4:10-4:30	孔德亮 (中国石油大学)	基于语料库的学术英语教材语域变异多维分析	
10	4:30-4:50	赵鹏飞 (香港教育大学)	Textual organization of MA theses abstracts by Chinese and English students of applied linguistics: Features and failures	
11	4:50-5:10	周海媚 (华南师范大学)	Interdisciplinary Uniformity and Variation of Lexical Bundles in Academic Writing	



**分论坛 2：基于语料库的话语分析（腾讯会议：992709286；密码：435435）**

**主持人：**钱毓芳教授 浙江工商大学；郇昌鹏副教授 上海交通大学

序号	时间	报告人	题目	主持人
1	1:30-1:50	李均桥 （华南理工大学）	语境重构与国家形象建构：基于语料库的新冠疫情报道研究	钱毓芳教授
2	1:50-2:10	李涛 （上海海洋大学）	语料库历时话语分析路径的中国政治概念英译规范嬗变考察	
3	2:10-2:30	李卓然 （上海外国语大学）	Critical Discourse Analysis of Reporting Practices in English News Reports on Public Health Event in China and US	
4	2:30-2:50	毛浚语 （北京外国语大学）	基于语料库的冬奥新闻文本批判性话语分析——以阿拉伯和美国主流媒体为例	
5	2:50-3:10	王佳宇 赵亚茹 （华中师范大学）	The Newsworthiness of Gu Ailing: A Comparative Critical Discourse Analysis of Chinese and American Media Reports	
6	3:10-3:30	王志威 （爱丁堡大学）	“中国人在线”——以互联网为媒介的中国国族认同的话语（再）生产	郇昌鹏副教授
7	3:30-3:50	吴祖昊 毕玉德 （复旦大学）	韩国主流媒体新冠疫苗舆情主题演化分析	
8	3:50-4:10	张鹏 （西华大学）	《华尔街日报》中华为企业形象的“话语-历史”分析	
9	4:10-4:30	吴佳媚 （上海外国语大学）	基于语料库的中美新闻语篇语法隐喻研究——以CNN和CGTN上海疫情报导为例	
10	4:30-4:50	钱玉彬 钟启芬 程子涵 （中国科学院大学）	A stylometric approach to the interdiscursivity of professional practice	
11	4:50-5:10	张星霞 （韩国汉阳大学）	A Study of China's Diplomatic Neologisms: The Language Situation in China from 2010-2020	



**分论坛 3：**基于语料库的语言对比、学习者语言研究（腾讯会议：631675611；  
密码：637053）

**主持人：**邵斌教授 浙江大学；吴淑琼教授 四川外国语大学

序号	时间	报告人	题目	主持人
1	1:30-1:50	房印杰 (北京邮电大学)	中国学生英语系词 be 过度生成句的准历史构式分析	邵斌教授
2	1:50-2:10	陈逸南 (华南师范大学)	What syntactic features predict L2 writing quality? A corpus-based study of letter compositions by different proficiency level Chinese learners of English	
3	2:10-2:30	吴淑琼 (四川外国语大学)	词汇语义的语料库量化研究	
4	2:30-2:50	李悦洋 (华东师范大学)	A Corpus-based Analysis of Chinese Advanced EFL Learners' Use of the Conjunction As in Argumentative Writing	
5	2:50-3:10	何鸿鹰 (贵州师范大学)	基于语料库对粤方言情态副词“好彩”的语义功能研究	
6	3:10-3:30	邵斌 陈泽南 (浙江大学)	20 世纪以来语境研究范式的变迁与进展	吴淑琼教授
7	3:30-3:50	邢富坤 (浙江外国语学院) 徐佳宁 (青岛大学)	汉英机器翻译的小句复合体转换能力调查与分析	
8	3:50-4:10	黄静茹 (华南师范大学)	基于语料库分析的商务英语教材中思辨能力培养探究	
9	4:10-4:30	高歌 (北京交通大学) 卫乃兴 (北京航空航天大学)	Revisiting semantic prosody in the interface between English and Chinese	
10	4:30-4:50	康锐江 (北京外国语大学)	基于语料库的对外汉语教材中国文化符号研究	
11	4:50-5:10	李想 (北京师范大学)	基于语料库的英汉“愤怒”类情感隐喻词汇对比研究	



**分论坛 4：基于语料库的翻译研究**（腾讯会议：325325560； 密码：908388）

**主持人：**黄立波教授 西安外国语大学；庞双子长聘副教授 上海交通大学

序号	时间	报告人	题目	主持人
1	8:30-8:50	崔舒洋 (北京大学)	基于语料库的经济政策类中央文献译本词汇特征研究 ——以《政府工作报告》(2016—2021 年)阿译本为例	黄立波教授
2	8:50-9:10	郭晓润 (郑州大学)	基于语料库的《楚辞·离骚》英译本翻译风格对比研究——以霍克斯、杨宪益戴乃迭、许渊冲译本为例	
3	9:10-9:30	刘昌 (青岛大学)	基于语料库的《论语》英译本语域变异多维分析	
4	9:30-9:50	陈洁玮 (马来亚大学)	新时期的中国“仁”形象建构：一项基于语料库的习总书记新年贺词翻译研究	
5	9:50-10:10	王月 (郑州大学)	MDA 模型驱动下《道德经》英译本语域变异的多维度分析	
6	10:10-10:30	吴诚辉 (对外经济贸易大学)	目的论视角下的基于语料库的《红楼梦》英文全译本语域变异多维分析	庞双子长聘副教授
7	10:30-10:50	杨菲 (浙江工商大学)	词语搭配视角下的“境界”英译研究	
8	10:50-11:10	应小寒 刘媛 (西安科技大学)	基于语料库的《政府工作报告》翻译语言特征研究	
9	11:10-11:30	张越 (香港中文大学)	A Corpus-Based Study on the English Translation of “Shuodao” in Honglouloumeng	
10	11:30-11:50	杨婷玉 (新疆大学)	试析基于语料库的维吾尔语动宾短语的汉译策略	



分论坛 5：基于语料库的汉语研究（腾讯会议：580624065；密码：803387）

主持人：郭鸿杰教授 上海财经大学；郭曙纶副教授 上海交通大学

序号	时间	报告人	题目	主持人
1	8:30-8:50	韩一 （中国社会科学院大学）	汉语母语者与学习者学术写作元话语句干语料库对比研究——以汉语国际教育专业硕士学位论文为例	郭鸿杰教授
2	8:50-9:10	董紫薇 （阜阳师范大学）	“破”类词语同一性问题及其历时演变探究	
3	9:10-9:30	郭曙纶 崔雪涵 （上海交通大学）	《人民日报》70 年“发展”左邻词分析	
4	9:30-9:50	李占 （北京信息科技大学）	“就你”的话语功能探析	
5	9:50-10:10	刘易 （北京师范大学）	基于 HSK 动态作文语料库的汉语兼类词计量研究	
6	10:10-10:30	刘娜 （北京航空航天大学）	Comparing Analytic Causative Constructions with shǐ 使 and ràng 让 in Mandarin	郭曙纶副教授
7	10:30-10:50	刘运锋 （安徽工程大学）	局部语法视角下的修辞研究拓展——《诗经》中“不我 V”为例	
8	10:50-11:10	杨雅琴 （北京语言大学）	基于语料库的“赶快、赶忙”的多角度辨析	
9	11:10-11:30	张乐滔 （北京外国语大学）	基于多因素分析法的现代汉语假设连词研究	
10	11:30-11:50	钟子龙 （北京外国语大学）	基于语料库的《政府工作报告》中主题词“发展”及其搭配研究	



**分论坛 6：语料库文体学研究**（腾讯会议：471334614；密码：456653）

**主持人：**孙海燕 教授 河南师范大学；李鑫副教授 上海交通大学

序号	时间	报告人	题目	主持人
1	8:30-8:50	刘抒悦 （北京航空航天大学）	A Linguistic-Cognitive Approach to Mind Style of Characters in Copenhagen	孙海燕 教授
2	8:50-9:10	任芮奇 （上海交通大学）	从语言功能视角试探文学文本文体学研究中的词丛分类法	
3	9:10-9:30	宋亚军 （浙江工商大学）	The Research of Readers Emotional Meaning of Sun Tzu's The Art of War	
4	9:30-9:50	于露 （郑州大学硕士）	基于语料库对《喜福会》文体特征及其中国形象研究	
5	9:50-10:10	张淑月 （西南大学硕士）	Attitudinal Resources and Characterization of Female Images in The Moon and Sixpence: A Corpus-based Appraisal Stylistic Approach	
6	10:10-10:30	钟洁 （上海海事大学）	A corpus-based analysis of gender stereotypes in Harry Potter	李鑫 副教授
7	10:30-10:50	周兵 （兰州大学）	数字人文视域下的西安城市想象与文化记忆	
8	10:50-11:10	李梦雪 （江南大学）	基于多模态语料库的民国时装杂志新探： 蒋兆和、《美的装束》与 1927 上海潮流切片	
9	11:10-11:30	彭小芳 （陕西师范大学）	基于语料库的汉语通用阅读教材中的中国形象研究——以《发展汉语》为例	
10	11:30-11:50	韩焜镁 （新加坡国立大学）	A corpus-linguistic approach towards language changes in normal aging	



**分论坛 7：语料库技术研究**（腾讯会议：466890931；密码：360779）

**主持人：**陆军教授 扬州大学；管新潮副教授 上海交通大学

序号	时间	报告人	题目	主持人
1	8:30-8:50	董书敏 (中央民族大学)	多语汉外新闻可比语料库构建方法研究——以汉-英-维-藏新闻可比语料库为例	陆军教授
2	8:50-9:10	林立 (新加坡国立大学)	Dispersion as a Measure of Grammatical Variation	
3	9:10-9:30	刘征 (中央民族大学)	面向中文突发公共事件主题的新闻语料库构建方法研究	
4	9:30-9:50	吴嘉静 (北京师范大学)	基于戏剧剧本的多模态语料库建设研究——以曹禺四大名剧为例	
5	9:50-10:10	张元卉 叶丽莺 (北京理工大学)	日语二语教学与习得的构式库建设——以教材中的构式为对象	
6	10:10-10:30	赵冲 许家金 (北京外国语大学)	BERT-assisted sense identification: The case of over	管新潮副教授
7	10:30-10:50	赵曦 肖巍 (重庆大学)	语料库视角下的中国哲学社会科学热门主题挖掘和发展历程分析	
8	10:50-11:10	袁咏雪 (江南大学)	基于 NDL 模型的与格交替多因素分析——以 Send 类为例	
9	11:10-11:30	李元科 (华南师范大学)	Synergistic effects of multiword sequences structure, function, frequency and association on raters' evaluations of essay quality	
10	11:30-11:50	李维静 (北京外国语大学)	A multimodal corpus analysis of the sub-genre variations of Western paintings	



### 三、 主旨发言

#### 主旨发言（一）



Professor **Wei Naixing** is currently honorary Chairman of Corpus Linguistics Society of China and Chairman of Academic Committee of the Foreign Languages School at Beihang University. His major research interests cover Corpus Linguistics, Learner Interlanguage Studies, Academic Discourse Analysis and Contrastive Studies of English and Chinese, for which he has published profusely, along with undertaking numerous state-sponsored research projects. He has also

served successively as Professor of English at Shanghai Jiao Tong University, Visiting Professor at Hong Kong City University and Hong Kong Polytechnic University, and Adjunct Professor at other key universities across Chinese Mainland, supervising/training corpus linguists of the younger generation and helping push forward corpus linguistics.

#### 题目：**Corpus Approaches to Discourse Studies: methodologies, Challenges and Prospects**

摘要：Corpus linguistics has offered a wealth of analytical instruments and techniques to discourse analysis, particularly, to the special field of Critical Discourse analysis. The synergy of the two areas of research has, therefore, opened up a new avenue of research in discourse studies, with unique methodological and theoretical values. In this talk, I set out to give a review of the methodologies of this avenue of research. I classify the research activities into three major types, namely, Corpus-based Discourse Studies (CBDS), Corpus-assisted Discourse Studies (CADS) and Corpus-informed Discourse Studies (CIDS), which are seen as lying on a continuum from corpus linguistics to discourse analysis, with CBDS and CIDS on two ends bordering corpus linguistics and discourse analysis respectively and CADS in between. Distinctive methodological features and underlying theoretical tenets of CBDS, CADS and CIDS will be spelt out and strengths and weaknesses of each approach commented on along the way. I will also discuss the challenges posed by the current accelerated boom of new information technologies for corpus-based discourse studies and the future prospects for development.



## 主旨发言（二）



**李文中**，浙江工商大学外语学院教授，博士生导师，1999 年获得上海交通大学博士学位，国家社科基金项目评审及鉴定专家。研究方向为语料库语言学、语料库与翻译研究、语料库与数据科学应用研究。已在国内 CSSCI 期刊发表学术论文 30 多篇，出版专著 2 部，译著 5 部，编著 2 部；已主持完成国家社科基金项目 3 项，主持完成教育部项目 2 项，主持完成省部级科研项目多项，参与完成多项国家社科基金项目。目前主持在研国家社科基金后期资助

项目 1 项（《语料库与数据科学应用》），以及浙江省哲社项目 1 项。兼任中国英汉语言比较研究会语料库语言学专业分会副会长、英国学术期刊 *IJCL*、中国学术期刊《语料库语言学》《中国外语教育研究》《中国应用语言学》等编委和审稿人。

### 题目：Textual Data Science for Corpus Linguistic Exploration of Textual Objects and Their Paraphrases

**摘要：**Textual (Linguistic) Data Science is defined as applying data science, which as an interdisciplinary field “uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data” (Wikipedia 2022), to the analysis of corpus documents, “using exploratory multivariate statistical methods” (Bécue-Bertaut 2018: xv; see also Treiblmaier 2021) to interpret meaning, extract information and gain insights from texts. While a corpus is genuinely recognized as a big language data source, applying data science to corpus data exploration is yet within the full reach of corpus linguists, still much less known to language researchers at large. Presented in this paper are various approaches to processing, visualizing, and exploring textual data, in order to obtain intratextual paraphrases around a set of closely-related textual objects, together with a preliminary discussion on their validation and justification. The term “textual objects” is coined following Teubert’s “discourse objects” to indicate any meaningful entity in a given text. Starting from its first occurrence in the text, its meaning arises from the co-selection with its cotext, and then such meaning prospects on-going paraphrases as the text unfolds itself (Sinclair 2004: 13). The interrelationship between different textual objects can be effectively explored using data science technology. My tentative conclusion is that corpus linguistics in full integration with data science better reveals the hidden patterns in textual meaning paraphrases, which helps us made informed decisions as what to look at and where to start from. Moreover, it is well advised for corpus researchers to be well equipped with the necessary data processing and analyzing skills, while being fully aware of the various data structures and their statistical as well exploratory applications. Finally, textual data science helps us to explore the text, but we always need to go back to the text and read to observe how meaning is created as the text flows.



### 主旨发言（三）



**邓耀臣**，大连外国语大学教授、博士生导师，《外语与外语教学》主编，享受国务院政府特殊津贴专家。研究方向为语料库语言学理论及应用、计算术语学。主持完成、在研国家社科基金一般项目各 1 项，主持在研国家社科基金重大项目子课题 1 项，参与国家社科基金重大项目，国家社科基金、教育部社科基金一般项目多项。在 *Applied Linguistics*、*International Journal of Corpus Linguistics*、*Journal of Quantitative Linguistics*、《外国语》、《中国外语》、《外语与外语教学》等国内外外语类权威期刊发表论文 40 余篇，主编出版学术专著 3 部。

题目：数字人文研究述评：热点领域与研究趋势

### 主旨发言（四）



**Michaela Mahlberg** is Professor of Corpus Linguistics at the University of Birmingham, UK. Before she came to Birmingham, she was Professor of English Language and Linguistics at the University of Nottingham. She previously held positions at the University of Liverpool, Liverpool Hope University College and the University of Bari. Michaela is the editor of the *International Journal of Corpus Linguistics* (John Benjamins) and together with Gavin Brookes she edits the book series *Corpus and Discourse* (Bloomsbury). Her publications include the monographs *Corpus Stylistics* and *Dickens's*

*Fiction* (Routledge, 2013), *English General Nouns: A Corpus Theoretical Approach* (John Benjamins, 2005) and *Text, Discourse and Corpora. Theory and Analysis* (Continuum, 2007, co-authored with M. Hoey, M. Stubbs & W. Teubert). Over the past couple of years, Michaela has been leading the development of the CLiC web application. Michaela is a Fellow at the Alan Turing Institute in London. She is the Vice President of the Dickens Society and the host of the podcast “Life and Language”.

题目：(Corpus) linguistics in the digital age

摘要：When corpus linguistics started off, the innovation it brought to linguistics



crucially depended on computational tools and methods. Since the early days, methods to store, collect, and analyze language data have developed significantly. Many areas of linguistics have now seen a ‘corpus’ or a ‘computational turn’. Beyond linguistics, language data is increasingly studied in a wide range of subjects and areas of application – from literature to political sciences. As a result, we sometimes see parallel developments across different fields. At the same time, there is ever-growing interest in cross-disciplinary exchange. The rapid development of computational methods and innovations in data science and AI, for instance, have much to offer to the study of language. We are at a point now, where we need to be clear about the foundations but also the future directions of corpus linguistics. Initially, corpus linguistics brought the computational into linguistics. Now, it is the task of corpus linguistics to keep the focus on the language in a world of data. This paper will outline where I would like to see corpus linguistics in the digital age.

### 主旨发言（五）



**Wolfgang Teubert** is Professor of Corpus Linguistics at the University of Birmingham, UK. He is the founding editor of *International Journal of Corpus Linguistics* (John Benjamins), the co-editor of *Studies in Corpus and Discourse Series* (Continuum) and the consulting editor of *Studies in Corpus Linguistics* (John Benjamins). His major research interests cover corpus linguistics, discourse analysis. His publications include the monographs *Meaning, Discourse and Society*, *Text Corpora and Multilingual*

*Lexicography*, *Corpus Linguistics: A Short Introduction*, *Corpus Linguistics: Critical Concepts in Linguistics*. He previously held visiting professorships at Communication University of China, Yanshan University, Henan Normal University, etc. He has supervised/trained many corpus linguists of younger generation and made significant contribution to the development of corpus linguistics.

### 题目：Why Discourse is Not an Object of Science

**摘要：** In spite of what many linguists believe, linguistics is not a science. It does not describe and explain the natural world. Its object is discourse, this entirety of utterances that have been made and are being made, as long as people keep talking to each other. By talking to each other, they exchange and share content. What they say has meaning. But while corpus linguistics has developed tools to calculate the collocation profiles of lexical items, meaning is not calculable.

Language signs, i.e. words or phrases, mean. But their meaning is not fixed; it is arbitrary. Some people say meaning is someone’s mental or neuronal representation of



a word. Others have claimed the ‘true’ meaning of a word is what the ‘thing’ to which it refers ‘really’ is. For me, however, language is what happens between people. This is why meaning only exists in discourse. Take a word like *depression*. When you ask people what depression is (co-extensive to what the word *depression* means) you get a lot of different answers. Discourse speaks in many voices. The meaning of depression is everything that has been said about it, and there is hardly a common denominator. So every new paraphrase will add to the meaning of this word. We may not know if depression is something that exists in the ‘real’ world out there, in discourse-independent reality. But, like the whole discursively constructed reality, this concept has a strong impact on our lives.

As I see it, our most important task as linguists is to find out what things said mean. This means we first have to assemble and select the relevant discursal evidence, to order it, and then to make sense of it. The first two steps are methodical, and we, the peer community, will easily agree on them. This is the playing field of corpus linguistics. But our task does not end there. For the third step, for the interpretation of the evidence, there is not much of a method interpreters have to follow. In principle, they are free to speak their mind. If we ask a dozen people to tell us what this discursal evidence of *depression* means to them, we will get twelve different answers. There is no ‘true’ and no final interpretation. But good interpretation needs guidance, and guidance is what corpus linguistics can provide.

In discourse, we, the discourse participants, create the world we share by talking about it. Discourse is not a thing we find in the ‘real’ world. Science is there to explain what happens in the natural world. It can explain it in terms of laws of nature. But discourse is not a natural object, it is an artefact. It is a human fabrication. And it does not obey any laws. It is discourse that makes us self-aware, makes us feel, gives rise to our attitudes and opinions, forms our mindset. Discourse frames how we perceive the world around us. Discourse is what shapes us. My point is: Corpus linguistics must re-invent itself. It must come to terms with interpretation. Our new focus must be meaningful discourse analysis.