

篇章分析——对汉语、法语 指称链条分布规律的 实证研究及其标注

王秀丽 胡 庆 编著



北京语言大学出版社
BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

篇章分析——对汉语、法语 指称链条分布规律的 实证研究及其标注

王秀丽 胡 庆 编著



北京语言大学出版社
BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

© 2018 北京语言大学出版社, 社图号 18100

图书在版编目 (CIP) 数据

篇章分析 : 对汉语、法语指称链条分布规律的实证研究及其标注 / 王秀丽, 胡庆编著. — 北京 : 北京语言大学出版社, 2018.8

ISBN 978-7-5619-5256-6

I. ①篇… II. ①王… ②胡… III. ①法语—对比研究—汉语 IV. ①H32 ②H1

中国版本图书馆 CIP 数据核字 (2018) 第 191699 号

篇章分析——对汉语、法语指称链条分布规律的实证研究及其标注
PIANZHANG FENXI——DUI HANYU、FAYU ZHICHENG LIANTIAO FENBU GUILÜ DE
SHIZHENG YANJIU JI QI BIAOZHU

责任编辑: 吴 硕

责任印制: 陈 辉

出版发行: 北京语言大学出版社

社 址: 北京市海淀区学院路 15 号, 100083

网 址: www.blcup.com

电子信箱: service@blcup.com

电 话: 编辑部 8610-82303392

发 行 部 8610-82303650/3591/3648 (国内)

8610-82303365/3080/3668 (海外)

读者服务部 8610-82303653

网 购 咨 询 8610-82303908

印 刷: 北京建宏印刷有限公司

版 次: 2018 年 8 月第 1 版

印 次: 2018 年 8 月第 1 次印刷

开 本: 787 毫米 × 1092 毫米 1/16

印 张: 13.5

字 数: 272 千字

定 价: 62.00 元

PRINTED IN CHINA

中央高校基本科研业务费专项资金资助

2016 年北京语言大学校级重大基础研究专项（项目标号：16ZDJ06）研究成果

2017 年中法两国教育部“蔡元培”交流合作项目

——北京语言大学法语系与斯特拉斯堡大学 LiLPa 语言学实验室合作项目团队共同编写

◎ 前 言 ◎

时光荏苒，人过中年，芳华已过，提笔写序，已不笃定。放慢脚步，放松心情，回首望去，知识的宝库徜徉过，但不知道是否拾得麦穗儿；而如今，更加关注的是与我当初有同样激情与梦想的后生。所以，说来也不怪，我的人生的圈子如同大多数老师一样，转回初衷，教书育人，转回好友，三五同学，刮摩淬励，衣钵相传。

本书编撰的第一个目的就是提携北京语言大学法语系的年轻教师、博士生和硕士生。他们虔诚地在法语语言学的园地中耕耘，理应让他们有所收获。我虽不能点石成金，但至少毫不吝啬地给予了知识与方法，读者会从他们的论文和译文中尽窥其能。

本书编撰的第二个目的追踪篇章语言学国内外前沿。我们不是弄潮儿，但也不放弃与第一方阵保持近距离的勇气。鉴于此，本书的内容主要呈现了篇章语言学与计算语言学结合的动态。中国社会科学院语言研究所的老友方梅、徐赳赳研究员和法国篇章语言学家 Catherine Schnedecker、Céline Guillot、Frédéric Landragin 和 Sophie Prévers，亦或老友，亦或同学，分别为中法两国新一代篇章分析的领军人物，承蒙赐稿或承让版权，使得我们从老一辈那里学到源自西方的篇章语言学能够在遥远的东方薪火相传。

本书编撰的第三个目的是为组建北语的法语语言学团队搭建平台。几年来，在学校和外国语学部的支持下，已经有 6 名法语语言学博士、6 名法语语言学博士生、24 名硕士生落户法语系，其未来的发展潜力，不应低估。

最后要感谢北京语言大学科研处、北京语言大学出版社对本书所付出的努力与支持。感谢中法教育部“蔡元培”项目所提供的与斯特拉斯堡大学 LiLPa 语言学实验室同行切磋的机会。一本书从知识积累与沉淀到与读者见面，虽前后历经十余载，但总会存在这样那样的疏忽，敬请读者批评指正。

王秀丽

2018 年春

◎ 目 录 ◎

汉语、法语篇章衔接研究的新动态·····	王秀丽	1
汉语叙事语体中零形主语小句的句法整合功能·····	方 梅	15
新媒体篇章的词频分析——以 QQ 网聊为例 ·····	徐赳赳	37
汉法指称链条平行语料库的标注方法和原则·····	胡霄钦	43
百科文体中指称链条的汉法比较研究·····	郭 畅	53
Research Articles From <i>Plos Biology</i> : A Textual Data Analysis ·····	Bruno Oberle	60
对汉语语法名词回指限定成分的对比研究		
——以《包法利夫人》平行语料为例·····	王鹏飞	74
法文商务信函的语篇衔接特征·····	朱佳琪 田俊雷	83
文学文本中标点符号与篇章衔接手段的相互作用·····	胡 庆	90
对汉语、法语指称链条的对比研究		
——以加缪《局外人》中的补语从句为例·····	徐世闻	95
对汉法零回指的篇章对比分析·····	孙晨晨	103
汉法第三人称代词回指的对比分析		
——以《约翰·克里斯朵夫》及其译本的平行语料库为例·····	王 晶	114
汉法应用文本导游词回指现象的对比研究·····	方赵春	121
篇章指称链条与文本体裁·····	严美馨等(译)	129
书面文本中共指链条的分析与标注程序·····	吕玉冬(译)	150
对指称和指称过渡的分析: 计算机辅助语言学的贡献·····	王鹏飞(译)	163
计算机辅助指称链条研究: 方法论问题及技术解决方案·····	田 宇(译)	176
一种针对语篇自动标注的跨语言分析方法:		
针对法语和保加利亚语的应用·····	沈思元(译)	195

汉语、法语篇章衔接研究的新动态

北京语言大学法语系教授 王秀丽

摘 要：本文以话题链条和指称链条为轴线，梳理了近年我国篇章分析的主要研究成果，以计算机辅助篇章研究为切入点，呈现了法国篇章分析的最新动态，从而把篇章分析的过去、现在和将来放在一条线上，把国内和国外、南派和北派放在一个共同的视域下。

关键词：篇章分析；话题链条；指称链条；机器辅助

1. 引 言

无论是篇章分析还是话语分析，真正的语言学意义上的研究发生在二十世纪八十年代初引进西方现代篇章语言学理论后，区别于本土研究汉语篇章语言问题过于重视写作技巧而没有集中在挖掘篇章分析规律的时期，分析单位从此转向超越句子的语言单位，即“超句实体”所进行的各种“篇章分析 (analyse textuelle)”研究。真正意义的篇章分析更倾向于笔语分析，主要含有两个分析领域：话题链条分析（廖秋忠，1986；王秀丽，2008）和指称链条分析（陈平、徐赓赓、徐丹、方梅、许余龙）。前者的研究对象是那些语义较弱、组织篇章结构的表达方式，即一些列位于句首、语义弱化、引出各种话语范围框架、主要起承上启下作用的副词性成分，也就是所谓的“下走 (descendant)”；后者研究的有各种回指构成的指称链条，即植入客观实体的命名性成分，也就是“上走 (montant)”。

从上述两方面来看，具有现代意义的严格的语言学意义上的篇章研究，从国际通用的术语开始，是从外语界兴起的。这两方面的研究的中间力量及其理论基础、分析方法可看出端倪。由于篇幅问题，本文将着重介绍一下从二十世纪八九十年代开始的指称链条的分析，更确切地说是各种回指研究，还不是真正意义的“链条”的研究：

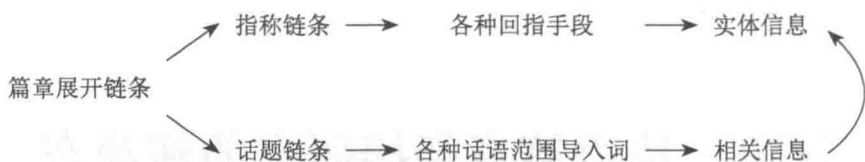


图1 篇章分析中的两个链条

我国外语学界对篇章的术语引进和应用到汉语的篇章结构研究也主要集中在以上这两个方面。

2. 对话题链条的研究

廖秋忠先生毕业于美国加州大学伯克利分校，1978年回国，1983年起在《中国语文》连续发表文章，这些文章后来均收录在廖先生去世后的纪念文集《廖秋忠文集》中，每篇都具有开创性的术语引进和方法引进的双重开拓意义。如：《现代汉语篇章中空间和时间的参考点》（《中国语文》1983年第4期）、《篇章中的框—根关系与所指的确定》（《语法研究与探索》1985年第3集）、《现代汉语篇章中指同的表达》（《中国语文》1986年第2期）、《现代汉语篇章中的连接成分》（《中国语文》1986年第6期）、《篇章中的管界问题》（《中国语文》1987年第4期）、《篇章中的论证结构》（《语言教学与研究》1988年第1期）、《物体部件描写的顺序》（《语言研究》1988年第2期）、《空间方位词和方位参考点》（《中国语文》1989年第1期）。

王秀丽老师博士毕业于索邦大学，师从法国篇章语言学家 Michel Charolles，2008 出版《篇章分析——汉法话语范围导入词对比研究》一书。全书共 10 章，主要分析了位于左侧句首的话题信息是如何作为篇章地图，引领局部三个连续句的篇章语法形成篇章的。从篇章性出发，落实到具体的语言与表达方式和形式，作为语句展开的动态推进器，左移成分在助力指称链条的同时，也形成了话题链条，对事件实体的状物性转化为衔接性，为语言编码增加了润滑剂。同年，作者的《话语范围导入词——对存在句的篇章分析》一文发表于《外语教学与研究》（2008 年第 5 期），首次以汉法存在句对比分析为出发点，从篇章的组织结构角度对传统的位于句子左侧的分离的地点状语专门进行了描写分析，给出了存在句在信息上与上

下句组成的话题链条及其衔接作用和认知功能,开拓了基于语料分析的篇章视域下的语言事实对比分析,有助于认清两种语言不同的实现衔接方式。

3. 对指称链条的研究

3.1 北派研究

我国较早的一批篇章分析研究者都有着扎实的外语功底,或本身就是英语、法语专业出身,又留学或工作在国外,比如陈平、徐赳赳、方梅、徐丹、钱敏汝等。因为这些人同时都曾经在北京的中国社会科学院语言研究所工作,我们把这一派权且称作北派。他们发表了最早的也是最具分量的篇章分析的论文和论著。其中以陈平的研究为早期代表,之后徐赳赳的研究紧紧跟上。

徐赳赳博士,中国社会科学院语言研究所研究员、当代语言学研究室副主任。Discourse Studies 杂志编委,《当代语言学》杂志编委。在《中国语文》《当代语言学》《外语教学与研究》《语言教学与研究》等杂志上发表论文 30 余篇,其中与指称链条分析相关的文章著有:《叙述文中“他”的话语分析》(《中国语文》1990 年第 5 期)、《叙述文中名词回指分析》(《语言教学与研究》1999 年第 4 期)、《现代汉语联想回指分析》(《中国语文》2005 年第 3 期)、《现代汉语篇章回指研究》(中国社会科学出版社 2003),另外还有其香港城市大学博士毕业论文 Noun-anaphora in Chinese Texts。《叙述文中“他”的论语分析》以汉语语料为出发点,主要从叙述文中的代词回指和名词回指出发进行了深入细致的分析、如名词回指的分类,“他”在篇章中分布的制约因素等。《现代汉语篇章四指研究》运用话语分析的理论和方法对现代汉语篇章回指现象进行了更为全面深入的探究,为建立研究汉语篇章回指的分析框架奠定了基础。《现代汉语联想回指分析》一文对“联想回指 (associative anaphora)”进行定义,——指的是没有明确先行词的回指现象,主要分析现代汉语联想名词回指这一篇章现象,区分了同指名词回指和联想名词回指,提出了汉语名词回指的分析框架,把联想回指分为上下义回指和关联回指。接着作者分析了汉语上下义回指和关联回指关系的不同类别,最后探讨了上下义回指和关联回指在篇章中不同的推进方式。

陈平的代表作《汉语零形回指的话语分析》(1987)选取了三种汉语常见体裁的语篇,

对其中的零形回指数量与分布进行分析和比较。从数量看，新闻类语篇和学术类语篇中的零形回指偏少，叙事语篇偏多。从位置分布看，新闻语篇最简单，叙事语篇最复杂。在共性方面，无论是零形回指的位置还是先行语位置，三种体裁都以主语为主，而且排在前三位的都是 S-S、T-S、O-S 位置关系的零形回指。其他位置关系的零形回指很少。这些共性对于进行汉语零形回指规则描写和计算机应用均具有指导意义。

王秀丽 2012 年发表的文章《篇章分析中的概述回指》中对法国篇章分析中“概述回指”理论的主要观点做了整体介绍和梳理，阐明了这一概念的篇章特性以及在篇章中使用的条件和范围，并着重分析了它改变篇章推进指向、加重认知成本的特性及其语言的表征策略。这一研究丰富了传统篇章分析中的回指方式研究，对深入回指研究有实际意义。其文中强调了概述回指有如下特征：回指的是前面的整个句子，而不局限在重提前句中某一个具体成分上；回指称对象不是前句的表层结构，而是一个思想，是对前面出现的成分或信息的概述性总结；其他回指手段的回指词主要是使用代词、名词，甚至是动词、副词，而概述回指采用的主要重提手段是概述性抽象名词或指示代词。本文还强调了概述回指的篇章性首先存在于被回指的句子和回指词之间本质上并不对称的关系：句法范畴上具有公开异质性，但是句法范畴上的公开异质性并不能掩饰语义范畴的同质性。其次，概述回指是一种结构关系，这种两个篇章片段之间的关系是明显不匀称的，非重复的、非传递的，其中回指词依赖于对被回指句的概述。概述回指的篇章性的第三个特征存在于它建立的回指关系不是一个简单的指称的同指关系，而是指称展开链条上的一种复杂的同指关系，建立在对前面各个指称因素的整合关系上，这种复杂关系有可能改变篇章推进方向，是一种比较复杂的建立在后台认知之上的手段。概述回指篇章性的第四个特征是它的认知层面，它建立在对前面信息进行认知处理的基础上。本文最后探讨了概述性回指与共指、重提和指示性回指的关系。

还有王志 1998 年发表的《篇章代词“它”用法探析》，对“它”在现代汉语篇章中的用法做了初步的探析。归纳了“它”的使用条件及篇章结构层次与“它”的使用情况。此外王灿龙的《“否则”的篇章衔接功能及其词性问题》（《汉语学习》2008 年 04 期），王蒨的《“如果”句的语篇分析》（广西民族大学硕士学位论文，2011 年），王振华的《连词“要是”的语义及篇章分析》（上海师范大学硕士学位论文，2015），方经民的《论汉语空间方位参照认知过程中的基本策略》（《中国语文》，1999 年第 01 期）对回指和衔接词的研究在分析广度和理

论深度上都有所拓展。此外方经民的另外一篇文章:《现代汉语第三人称代词指称及其语境制约——兼与日语第三人称代词比较》(《当代语言学》,2004年第03期),则考察了现代汉语第三人称代词的直指和回指所受到的语境制约。他通过分析指出,第三人称代词用于介绍型现场直指时,只受到信息原则的制约,但是用于议论型现场直指,则要受到礼貌原则的制约。

3.2 南派研究

南派研究以上海外国语大学许余龙教授及其博士朱玉生等进行的一系列回指范围的研究为主。许余龙老师承担过国家社科基金项目“英汉篇章回指的功能语用对比研究”(2005),其专著《篇章回指的功能语用探索——一项基于汉语民间故事和报刊语料的研究》(上海外语教育出版社,2004),从一个全新的角度探讨了回指研究,共分绪论、回指确认及其语言提示特征、语篇实体的主题性、指称对象的可及性及其编码、验证方法与数据库建设、篇章中的主题管理、回指确认原则与篇章回指、句内回指的形式句法和语用研究、回指确认原则与句内回指、结论等10章。在理论上,本书提出了一个以可及性和主题性为基本概念的回指确认原则,并在理论模式的建构过程中,一方面修正和发展了国外学者提出的某些观点,如篇章主题的定义、识别方法和分类,另一方面提出了一些全新的概念,如篇章中的期待主题和期待副主题。在研究方法上,本书采用理论探讨和实证研究相结合的方法,提出了一个具有可操作性的系统验证上述理论假设的方法,并据此建立自己的语料数据库,从实际篇章分析出发,在统一的理论框架中,全面系统地探讨各类回指词在篇章中的分布和指称特点,证明了上述篇章回指确认原则对自然语言中的篇章和句内回指现象都具有很强的解释力。

此外许余龙先生从语篇中“名词短语回指”“可及性”“主题性”等基本概念出发,通过如何确定回指语的指称对象和回指确认模式研究,努力证明了篇章理解过程与人的认知的关系。主要论文有:《英汉指称词语表达的可及性》(《外语教学与研究》,2000年第5期)、《语篇回指的认知语言学探索》(《外国语》,2002年第1期)、《语篇回指的认知语言学研究验证》(《外国语》,2003年第2期)、《汉语主从句间的回指问题》(《当代语言学》,2003年第2期)、《语篇回指实证研究中的数据库建设》(《外国语》,2005年第2期)、《从回指确认的角度看汉语叙述体篇章中的主题标示》(《当代语言学》,2005年第2期)、《话题引入与语篇回

指——一项基于民间故事语料的英汉对比研究》(《外语教学》, 2007年第6期)、《名词短语可及性与关系化——一项类型学视野下的英汉对比研究》(《外语教学与研究》2012年第5期)、《溯因推理与篇章回指理解》(《当代修辞学》, 2013年第1期)、《释几对与语篇回指相关的概念》(《外文研究》, 2013年第2期)、《汉语包孕句中反身代词的回指问题》(《外语与文化研究》第四辑, 2004)、《篇章回指理解的认知本质与推理基础》(上海外语教育出版社, 2008)、《英汉语下指的篇章语用功能分析》(《外语教学与研究》, 2007年第39卷, 第6期)。

上海外国语大学曹羽菲的博士论文《西班牙语定冠词回指功能研究: 汉西对比》, 主要研究西班牙语定冠词的回指功能, 包括直接回指和联想回指, 并与汉语中执行相应功能的名词回指语段进行比较, 分析异同。上海外国语大学语言研究所厚露莹的硕士学位论文《汉语朗读语篇回指现象的语音特性研究》(2014)通过语音实验和统计分析, 以重音级别、基频(F0)及停顿时长为研究参数, 从语音角度对汉语朗读语篇中的篇章回指现象进行了系统考察。其研究的主要发现有: (1) 对于零形回指, 分句首零形式后的边界音节更为重读, 两种类型的零形式相比, 有距零形回指形式之后的边界音节更为重读, 紧邻零形回指形式之前的边界音节更为重读, 就时长而言, 紧邻零形回指形式之前的停顿更短; (2) 对于代词回指, 第三人称代词回指形式“ta”的音高与时长表现与其句法位置密切相关, 且段首、句首及分句首边界位置“ta”的语音表现与篇章层级有关, “ta”在不同句法位置与周围音节具有不同的首选韵律模式, 信息结构与修辞表达为影响“ta”语音表现的重要因素; (3) 关于名词回指, 位于分句首的先行词与名词回指形式相比要重读, 且之前的停顿时长更长。基于结果, 本文提出了零形式底层音节重量投射的观点, 讨论了代词回指与篇章韵律, 并探讨了名词回指中重音与停顿在朗读语篇中的语用功能。本研究属于语音与句法及以上层面的接口研究, 可从语音角度丰富汉语篇章回指领域的研究、加深对汉语篇章回指现象的理解, 本文所得结果对汉语篇章韵律的研究具有促进作用, 并在语音合成领域具有应用价值。

4. 法国指称链条研究的机辅自动化倾向

法国的语言分析不是以研究所为地点分布, 而是以国家科研中心、各个大学组成的联合实验室为据点分布。篇章研究动态以三大实验室为核心的团队人员为中心, 如 Michel

Charolles、Sophie Prévost、Catherine Schnedecker、Georges Kleibert、Frédéric Landragin、Poileau Thierry、Céline Guillot。这三家语言学实验室由法国国家科学研究院（CNRS）支持的法国语料库研究三大实验室，即 Lattice 实验室、LiLPa 实验室和 ICAR 实验室，扮演着举足轻重的角色。Lattice 实验室全称为语言、语篇、信息处理与认知实验室（Langue, Textes, Traitements informatique, Cognition），位于法国巴黎，由巴黎第三大学和巴黎高等师范学院的专家学者合作进行研究，研究重点为语言可视化分析工具的开发与研究，代表性成果为标注语篇资料库（La Base EIOMSIT）和新闻语篇借词短语标注资料库（BSP）。LiLPA 实验室坐落于斯特拉斯堡，全称为语言学、语言与言语实验室（Linguistique, Langue, Parole），主要依托斯特拉斯堡大学的研究团队开展研究工作，其研究重点为语篇指称链条描写、实现方式与长度研究，以及多体裁文本指称链条的分布研究，代表性成果为自动标注工具 ANALEC 的开发与研究，以及新闻体裁语篇标注资料库。上述两大实验室的研究多集中于书面语资料库（即语料库）的开发、标注与研究，而 ICAR 实验室则将研究重点放在口语语料库的开发与研究上。该实验室位于里昂，为里昂高等师范学院和里昂第二大学的联合实验室，旨在对口语互动交际中的语言使用情况采取多维度分析研究，其代表性成果便是 CLAPI 多模态互动口语语料库（Corpus de Langue Parlé en Interaction）。为了方便计算机用户达到在海量信息中提取专门信息的最终目标，词汇单词的链接、衔接词的链接是相当重要的，尤其是从篇章信息层面出发的寻找更有意义。于是从法语各种体裁样本出发，法国研究在指称链条的分布规律上力图展现出自动统计、数据统计的特点，凸显分布的衔接不同以及指称格局焊接点的距离远近，从而整理出法语衔接策略的独特特性。

4.1 自动化研究

不难理解，篇章信息的自动处理与语篇的语言结构表征有很大关系，就是要找出那些衔接性的关联结构，找出节点，构建语篇结构模型（patterns linguistiques），旨在方便文本的知识网络和语义网络的建构，达到机器理解的目的。为此，为了让机器学会篇章结构，细致到无穷的统计数字一定是必要的系统化、信息化、自动化的前提。具体地说，篇章领域的自动化就是把指称链条、话语范围导入词都看作一种标记，首先用设计的软件标注出来，然后提取各方面文本信息，虽然是表层的，但足够与上下文、语义、知识网络区分。指称表达方式、

导入区域性话题，并通过各种重新命名表达，是作者设计结构、读者追寻理解话题的可信赖的迹象，法国巴黎高师和巴黎三大联合国家实验室 LATTICE 的标注软件 ANALEC，有能力进行这一领域的标注（虽然有一定的错误率）。以一篇法语语料莫泊桑的小说《Les Bijoux》的主要人物指称链条的手工标记和软件表征为例^①：

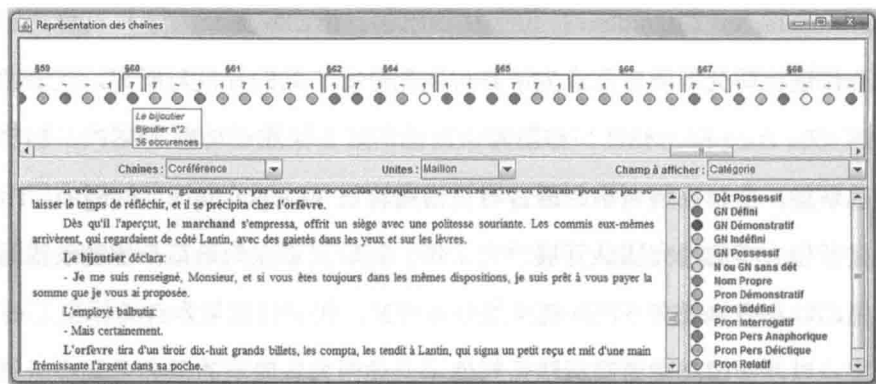


图2 对指称链条的手工标注

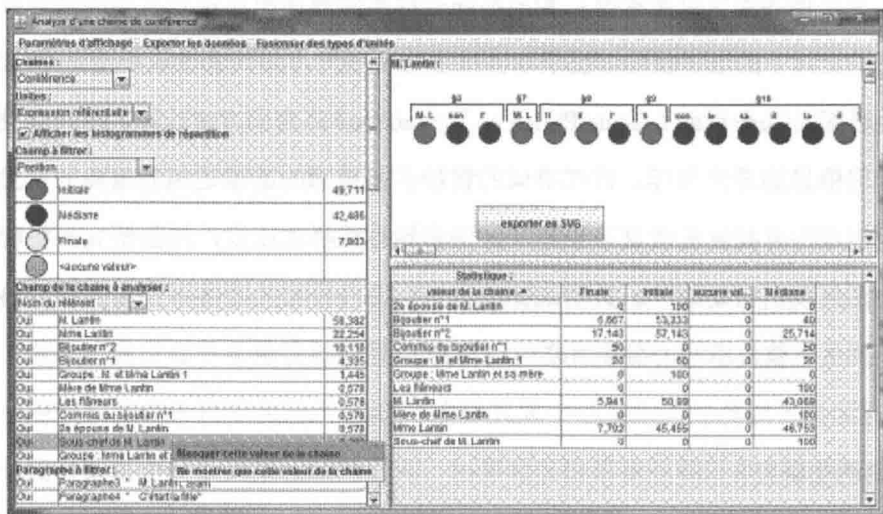


图3 ANALEC 对每个指称链条的分段呈现

① 本系列用图参照了法国国家科研中心巴黎高师和巴黎三大联合国家实验室 LATTICE 首席研究员 Frédéric Landragin 先生 2016 年在“国际篇章语料统计分析大会”（Les Journées internationales d'Analyse statistique des Données Textuelles）的发言《共指指称链条的研究和视觉化设计》（Conception d'un outil de visualisation et d'exploration de chaînes de coréférences），谨此致谢。

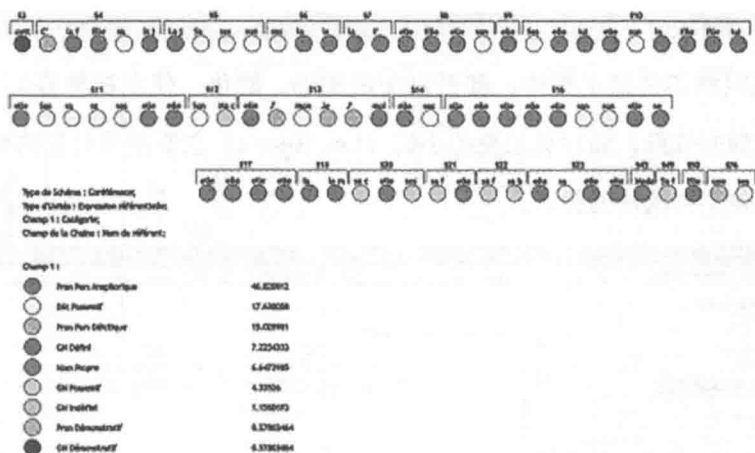


图4 计算机辅助表征指称链条的构成

4.2 数字模型化研究

质性分析与量性分析一直是篇章分析的研究特征，也是它的规律特征。追寻话题链条、指称链条也可以通过更可靠的数字标识和统计表现出来。在指称语言形式的研究分类基础上，对不同体裁的指称链条的形式和分布规律进行模型化，尤其是连接点的分类模型化，可以得出这些标注的指称语言表达方式对篇章形成的衔接作用，尤其是在不同语言中不同的分布图形、凸显、主题推进规律。模型化的前提应该是大规模多体裁语料库的建立，以便有足够数量的变体，实施对指称链条的组成，其中包括过渡、凸显质和量的分析，从简单的开始，逐步探索出自动辨识软件，实现自动标注和同步可视化，甚至是图形化和数字化试验。如图所示：



图5 指称链条的分布和整合

4.3 视觉化研究

不论是知识网络还是语义网络，都可以使用图形、颜色，使之在屏幕上更好地可视化，呈现纸板书达不到的效果。如对莫泊桑的小说《Les Bijoux》主要指称对象的识辨标识是这样呈现的，使用的是不同颜色的圆点：

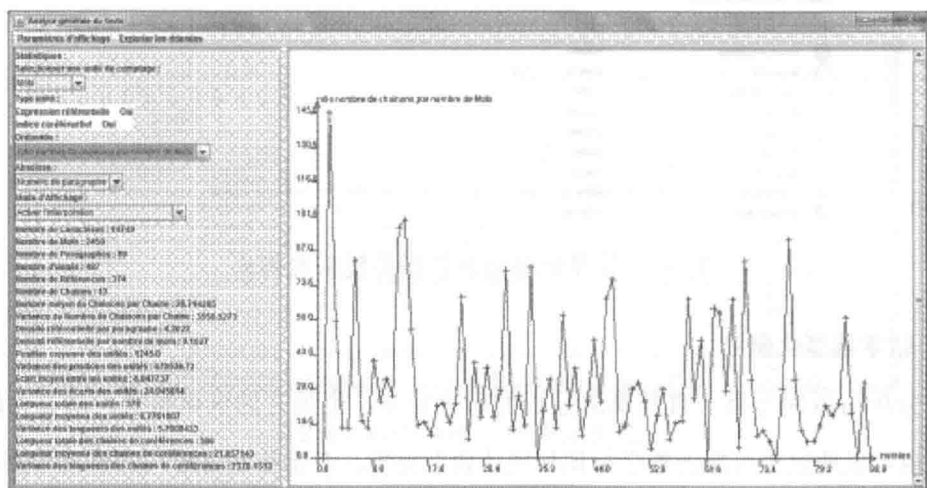


图6 ANALEC 表征的指称链条段密度比、指称数量词数量



图7 指称出现的视觉化表征