话语标记概貌分析与情感倾向探索

阚明刚 杨江 著

吉林文史出版社

国家社科基金青年项目“基于语义方法的汉语文本情感自动分析研究”(11CYY032)
河北省社会科学基金项目“英语同义词语体对比研究及平台建设”(HB14YY026)
2016 年度河北省社会科学发展研究课题 (201603010103)
华北理工大学博士科研启动项目

话语标记概貌分析与情感倾向探索

阚明刚　杨　江　著

图书在版编目（CIP）数据

话语标记概貌分析与情感倾向探索 / 阚明刚，杨江著．
-- 长春：吉林文史出版社，2016.9
ISBN 978-7-5472-3457-0

Ⅰ．①话… Ⅱ．①阚… ②杨… Ⅲ．①话语语言学
Ⅳ．① H0

中国版本图书馆 CIP 数据核字 (2016) 第 220891 号

话语标记概貌分析与情感倾向探索

HUAYU BIAOJI GAIMAO FENXI YU QINGGAN QINGXIANG TANSUO

著者 / 阚明刚　杨江
责任编辑 / 李相梅
责任校对 / 赵丹瑜
封面设计 / 西子
出版发行 / 吉林文史出版社有限责任公司
（长春市人民大街 4646 号，电话：0431-86037501）
www.jlws.com.cn
印刷 / 北京兴湘印务有限公司
出版日期 / 2017 年 3 月第 1 版　2017 年 3 月第 1 次印刷
开本 /710mm × 1000mm　1/16
字数 /200 千字
印张 /18
书号 /ISBN 978-7-5472-3457-0
定价 / 38.00 元

浩歌一曲酒千钟。男儿行处是，未要论穷通。

作者简介

阚明刚，男，汉族，现任教于华北理工大学外国语学院英语系，文学博士，毕业于中国传媒大学文学院，主攻语言学及应用语言学专业，中国人工智能学会会员。教学中主要负责讲授本科段《高级英语》《英汉翻译》《高级听力》《词汇学》等课程；研究生阶段的《研究生英语》以及《应用语言学》等课程。到目前为止，已在《计算机工程与应用》《语言教学与研究》等杂志上发表学术论文二十余篇，参编著作两部，出版学术专著一部，译著两部，合著专著一部，合译译著一部，独立主持完成教育部人文社科项目、河北省社科基金项目、河北省社科联项目各一项，现主持河北省社科联项目一项，获发明专利两项。

杨江，男，汉族，现任教于湖南科技大学外国语学院，文学博士，副教授，硕士生导师。毕业于中国传媒大学文学院，主攻语言学及应用语言学专业，中国人工智能学会会员。教学中主要负责讲授本科段《语言学概论》《语言信息处理》等课程；研究生阶段的《计算机辅助翻译》等课程。科研目前专注于语言主观性的计算、文本语义倾向的计算、文学内容计算和语言计量对比研究，侧重基于协同语言学的英汉语词汇语义对比分析。到目前为止，主持国家社科基金、省级社科基金、教育厅优秀青年基金等科研项目五项，在《当代语言学》《中文信息学报》等刊物上发表论文二十余篇，参编著作多部。

前　言

话语标记，即标记话语的话语，体现着人的思维过程、说话角度、说话方式、语篇构建的模式。汉语中有多少个话语标记、其功能状态如何都是需要研究的问题。同时，这些方面又是话语主观倾向性的体现。因此，在研究整体面貌的基础上，对话语标记进行情感分析就是研究的一个重要延伸方向。

本书首先对话语标记研究历史进行了追述，并对整体的研究意义和方法进行了阐述。接着构建语体语料库，从中提取全部规定的话语标记和实例。然后对话语标记系统进行整体对比分析和研究。这中间我们构建了功能标注体系，从数量上进行了统计。在对全部话语标记做了质和量的仔细研究后，我们又构建了主观倾向性体系，运用该体系进行了语料库的标注和计算实验。在测定其有效性后，我们对200个活跃的话语标记进行了主观倾向性分析，并以实例说明了分析过程。本书最后对研究内容做了总结，并对未来研究工作做了展望。

本书的第一章和最后一章，可以让读者了解话语标记研究的来龙去脉。这对理解本书对话语标记研究的角度和方法起到了导向作用。由于话语标记的复杂性，我们在第二章对性质和定义进行了仔细分析，并针对性地从计算语言学角度对本书的研究范围进行了限定。这是我们研究的出发点，也为我们的研究指明了方向。第三章是动手的第一步。本章中记录了语料库的建构原则和过程，并对前面的研究假设进行实验验证。结果表明：我们的设想是可行的。第四章就是对话语标记的全面提取，即看清整个话语标记体系，并对该体系进行分析研究。研究内容涉及到话语标记的语体分布、使用频率、功能类型、话语语篇位置等等。

看清话语标记概貌后，第五章就展开了对话语标记的主观倾向性研究。本章中，构建语言主观倾向性体系是第一部分的内容。主观情感的描述使用了六个维度，考虑了类别、形式、程度、模式、成分和关联。

在这个体系的基础上，本章第三部分建立了语义倾向语料库，为后续计算建立了工具箱系统。第六章是对主观性体系和语义倾向性计算的具体运用。为实现对话语标记主管倾向的研究，本章首先研究了基于句子规则的情感计算。第二部分是基于浅层篇章结构对评论文进行了情感倾向计算，指出了话语标记在主观倾向计算上各个层面的功能。第三部分是结合第四章的研究，运用主观倾向性体系，对200个活跃话语标记进行六维度倾向性分析。然后举例来说明分析过程和结果。这一章是话语标记整体研究的重要延伸，虽然只是初步探索，但也可以看出它们在语言体系中的重要性。

第七章是全书的总结。本章首先说明了研究的重要成果，然后分析了这些成果的应用价值，指出了研究的不足，并对进一步研究做出了展望。

本书的两位作者都是研究计算语言学的。使用语料库也是两位作者熟悉的研究手段和方法。不管是前期的话语标记系统的获得，还是后期主观倾向性的标注，都是在一定规模的语料库中进行的。这是本书的一个特色。另外，除了使用已有软件外，两位作者还自编了几个适用于本研究的软件，以解决必要的相关问题来推进我们的研究，比如话语标记提取软件、语体计算软件、评论文主观倾向性计算软件、词语新义自动发现软件等等。因此本书不只是语言类型的研究，还有工科的学科内容。这是本书的另外一个特色。

本书两位作者分担了不同章节的撰写，其中第一、二、三、四、七章由第一作者撰写，第五章由第二作者撰写，第六章由两位作者合作完成。

由于作者水平有限，在研究和撰写过程中，可能会存在着这样那样的问题，请各位专家学者不吝指正。

2016年7月

目　录

第一章 概 论

本章内容提要：本章对话语标记研究进行了综述，然后对研究的切入点、研究的意义和方法进行了简明介绍。

1.1 话语标记研究历程回顾[①]

话语标记是一种常见的语言现象。它在话语中使用的频率很高，形式较为固定，一般没有实际意义，是话语交际中的润滑剂。话语标记的研究是话语分析中的一部分，是在话语分析兴起之后近 20 年逐渐发展起来的，最近几年备受语言学界关注。国内外对话语标记的研究进程不十分相同，但却有很多相同的研究视角。国内外学界对话语标记从最初开始关注到目前多角度研究已经走过了 60 年，对话语标记的认识也从先前的“可有可无的语言成分”转变为“语言建构的重要手段”再到“主观虚化的结果”。在认识到话语标记重要性的同时，研究也呈跨领域、跨学科的势头。

1.1.1 国外研究回顾

从话语标记开始被留意到目前多角度研究经历了 60 年时间，我们将这段时间划分为四个阶段。

第一阶段：孕育期（50 年代初—70 年代中期）

1953 年，Randolph Quirk 在他的《随意的交谈——日常口语的一些特征》讲座中，首次提到口语中常出现一种无用而且毫无意义的成分，比如 you know，you see，Well 等，Quirk 把这些成分称之为“修饰语”，并认为对这些成分进行研究应该具有相当重要的价值。Quirk 的这一看法在当时并没有引起人们的注意。在同一时代，认识到话语标记的独特之处的还有 Fries。Fries 在其著作 *The Structure of English*（1952）中，没有像许多传统

① 本节内容参见：阚明刚. 话语标记研究综述［J］. 现代语文（语言研究版），2012，05：103－107.

语法作者那样简单地划分出一类“感叹词”，而是基于频率分布分析法编成更加详细的语法。Fries 将功能词划分为 15 类。其中 well，oh，now，why 频繁出现在“回答部分”（response utterance uni）的起始部位；更多情况下是在连续（continuing）交谈的句子开头处。（黄大网 2001）这是首次用话语概念对它们进行描述，但由于当时的理论限制，以后 20 年鲜见有人对这类话语成分进行研究。

第二阶段：萌芽期（70 年代中期—80 年代中期）

话语分析逐渐兴起的 20 世纪 70 年代，当学者把目光转向日常口语之时，特别是那些直接蕴含话语间关系的语言要素的时候，这些似乎毫无价值的“修饰语”才得到真正的关注，成为语言学家研究的热点对象。语言学家意识到，就是这些成分，才使得人类自身言谈显得自然并合乎语法，才可以把人类语言和机器语言分辨开来。但是，这一时期专家学者多是把这些“修饰语”当成话语分析的手段，因此没有形成专门的领域来研究。

具有语用学奠基之作的 *Pragmatics*（Levinson 1983）一书，提出了需要研究“用来标记某一话语与前面话语之间存在的某种关系”的成分这一课题。在随后的语言学发展过程中，人们对话语标记的语用地位和研究价值有了更深的认识。尤其是功能语言学、语用学、篇章语言学、认知语言学的渐次兴起，推动了话语分析向纵深发展。话语标记作为人们交际过程中的语用手段，已经不再是可有可无的装饰成分，而是非常重要的语言建构机制。

第三阶段：成长期（80 年代中期—90 年代末）

当认识到话语标记的重要性之后，许多学者都从自身的研究领域出发，对这一成分进行探讨。如：Polanyi and Scha（1983）的 *The Syntax of Discourse*，最初把话语标记命名为话语标记机制（discourse signaling devices），后来 Polanyi 运用语言的话语模型将框架内的基础话语单元分成两种类型：一种是携带命题内容的基本话语构成单元（the elementary discourse constituent unit），一种是不携带命题内容的话语操作符（discourse operators），即我们所说的话语标记；Quirk 等人（1985）的 *A Comprehensive Grammar of the English Language*，从语义角度看待话语标记，认为它们是语义连接词（semantic conjuncts）；Schiffrin（1987）的 *Discourse Markers* 是集大成之作，在该书中，Schiffrin 创立了局部连贯说（local coherence），认为话语标记是话语的语境坐标；Erman（1987）的 *Pragmatic Expressions in*

English: *A study of you know, you see and I mean in face – to – face conversation* 首次在考察大量对话实例的基础上对此类表达法进行系统描写的；Blakemore（1987，1992）的 *Semantic Constraints on Relevance* 和 *Understanding Utterances* 是运用关联理论对话语标记所做的研究，认为话语标记可以起到明示话语和（认知）语境关系的作用；Fraser（1990）的 *An Approach to Discourse Markers*，认为话语标记是当前基本信息和先前话语之间的序列关系的体现，每一个话语标记都有一个最小的语用核心意义；Redeker（1990，1991）的 *Ideational and Pragmatic Markers of Discourse Structure* 和 *Review article*: *Linguistic Markers of Discourse Structure*，是 Redeker 对话语标记更全面和细致的研究，她认为 Schiffrin 的谈话五层面并不能对语篇连贯起到平等的作用；Knott & Dale（1994）的 *Using Linguistic Phenomena to Motivate a Set of Coherence Relations*，讨论了对提示短语（cue phrases）收集和分类，角度采用的是修辞结构理论；Ostman（1995）在著作 *Pragmatic Particles Twenty Years After* 中提醒说，“现在对语用标记研究有升温势头，实际上此类研究是要回溯到 60 年代末和 70 年代初，当然更早些的里程碑式的文献有 Denniston 的 *The Greek Particles* 和 Arndt 的文章 ‘Modal Particles in Russian and German’，而真正的语用标记或语用小品词研究上的突破则是 Weydt 的 *Abtönungspartikel*，等等。这些学术论著，由于研究的角度不同，因此对整个话语标记的研究都做出了应有的贡献。其中，Schiffrin 的影响最大，后来学者也都采用该著作中对这种成分的命名，即话语标记，来进行研究。本研究也采用“话语标记”这一术语。

从 1986 年国际语用学杂志 *Journal of Pragmatics* 专刊登载话语标记系列研究文章开始到 1998 年该刊再次出专刊，话语标记研究逐渐成熟起来。这个时期也是话语标记本体研究的深化和发展。

这个时期对“话语标记”的不同称谓也是个很有意思的现象。冉永平（2000）对称谓进行了总结归纳，列出了包括萌芽期开始使用的 25 个不同的名称。我们按照出现时间对它们重新排序：语句联系语（sentence connectives）（Halliday & Hasan 1976）、暗示词（clue words）（Reichman 1978）、外加语标记（disjunct markers）（Jefferson 1978）、语义联系语（semantic connectives）（van Dijk 1979）、话语策略语（gambits）（Keller 1981）、逻辑联系语（logical connectors）（Celce – Murcia，et al. 1983）、话语标记手段（discourse signaling devices）（Polanyi & Scha 1983）、导语

(prefaces) (Stubbs 1983)、语用联系语 (pragmatic connectives) (Stubbs 1983; van Dijk 1985)、话语小品词 (discourse particles) (Schourup 1985)、语义联加语 (semantic conjuncts) (Quirk, et al. 1985)、语用标记手段 (pragmatic devices) (Vande Kopple 1985)、语用表达语 (pragmatic expressions) (B. Erman 1987)、语用构成语 (pragmatic formatives) (Fraser 1987)、句外连接语 (extrasentential links) (Fuentes 1987)、表意联系语 (phatic connectives) (Bazanella 1990)、话语操作语 (discourse operators) (Redeker 1991)、指示手段 (indicating devices) (Katriel & Dascal 1992)、超命题表达式 (hyperpropositional expressions) (Moon 1992)、提示短语 (cue phrase) (Hovy 1994; Knott & Dale 1994)、语用操作语 (pragmatic operators) (Ariel 1994)、提示词 (cue words) (Rouchota 1996)、会话常规语 (conversational routines) (Aijmer 1996)、语用标记语 (pragmatic markers) (Schiffrin 1987; Fraser 1988, 1990, 1997)、语用功能语 (pragmatic function words) (Risselada & Spooren 1998)。从这些名称上我们可以依稀看出话语分析的兴起和语用学诞生痕迹，也可以看出对话语标记研究的不同层面和不同取向。

第四阶段：扩张期（21 世纪初—）

国外话语标记的研究在二十世纪末期已经如火如荼地展开了，文章著述越来越多。进入21 世纪后，话语标记研究采用的角度和方法也越来越丰富。Fox Tree (2001) 的 *Listener's Uses of um and uh in Speech Comprehension* 和 Clark & Fox Tree (2002) 的 *Using uh and um in Spontaneous Speaking* 两篇文章，都是从认知心理学的角度对话语标记进行研究；Simone Müller (2005) 的 *Discourse Markers in Native and Non - native English Discourse*，探讨了话语标记在二语习得中的应用和语料库驱动的话语标记研究，并对 so, well, you know, like 四个话语标记进行了定量分析；Kerstin Fischer (2006) 的 *Approaches to Discourse Particles* 一书展现了话语标记研究在方法上的广泛性和丰富性，手段是通过问询不同背景的专家学者并请他们对一些关键问题给出自己的看法，如：定义、功能和模型框架等等；Mariam Urgelles - Col 博士（英国 Middlesex 大学的副讲师）出版了他的专著 *The Syntax and Semantics of Discourse Markers* (2010)，该书研究了话语标记的句法和语义。句法方面的讨论是在头驱动短语结构语法 (Head - driven Phrase Structure Grammar) 框架下进行的；由于话语标记是在语篇层面运作

的，需要有完整的理论，因此作者采用了分段语篇表征理论（Segmented Discourse Representation Theory）。作者还综述了话语标记研究，范围覆盖从话语分析角度到关联理论的应用再到计算语言学的方法等诸多领域。

同样的话语标记研究热潮很快扩散到其他语言。例如，Noriko O. Onodera（2004）的 *Japanese Discourse Markers*，从共时和历时角度对日语的连词“でも”和“だけど”、感叹词“ね”和“な”进行了研究，Schiffrin 认为作者创建了理论语言学和传统历时语言学未来研究的重要指导原则；Montserrat González（2004）的 *Pragmatic Markers in Oral Narrative: the Case of English and Catalan* 一书，首先探讨了叙事框架，然后对比了 Schiffrin 和 Redeker 在话语标记研究上的不同，分析了话语标记在叙事结构中的作用；Catherine E. Travis（2005）的 *Discourse Markers in Colombian Spanish: a study in polysemy* 从话语标记的多义性出发，提出了话语标记使用上的社会变体，并通过建立语料库的方式，在讨论了什么是话语标记和前人如 Schiffrin 和 Fraser 的研究等之后，分析了话语标记的韵律独立性、句法独立性、语义独立性，指明了话语标记的三个功能：构建主要发话人的话轮的功能、语境功能和互动功能。作者构建了自然语义元话语方法，并将该方法运用到话语标记研究之中，分析了西班牙语中的 bueno、O sea、Entonces 和 Pues，等等。

总之，对话语标记的研究从无到有、从弱到强，经历了一个比较漫长的时期。开始是有专家学者的注意，然后是随着话语分析和语用学的发展而诞生，接着是对话语标记理论上的深入探讨，最后发展为多角度、多模式、跨学科的研究。

1.1.2 国内研究回顾

我们对国内话语标记研究按名实隐显也划分成了四个阶段。

第一阶段：有实无名期（19 世纪末—20 世纪 70 年代末）

汉语中对话语标记这类成分的讨论早已有之，是语言学家长期讨论的对象。最早是马建忠（1983）在《马氏文通》中的分析。他把起连接作用的词称为“虚字”，并将连接词通称为“连字”。以后许多语言学家都从语法—语义角度对这些词语进行性质与分类上的研究。赵元任（2001）指出，汉语中的连词和介词或副词难以区分；连词的位置具有不确定性。他把连词分为四类：a. 介词性连词，如“跟”“和”“同”等；b. 连词的超