

语义密度、句法复杂度与学科知识编码

高彦梅¹ 周江平^{1,2}

(1. 北京大学, 北京, 100871; 2. 西华师范大学, 南充, 637002)

摘要: 语义密度和句法复杂度揭示学科知识编码属性和学科文化特征。本研究通过对比分析中国本科生、研究生、学者和本族语作者学术论文摘要中语义密度和句法复杂度的分布, 探索语义密度和句法复杂度作为学科知识编码水平测量手段的可行性。研究发现, 在不同学科水平的作者之间, 语义密度和句法复杂度分布具有明显差异性。三组较高水平作者语义密度显著高于本科生作者; 母语作者的句法复杂度显著高于研究生和本科生, 但与中国学者没有显著差异。语义密度反应知识浓缩程度, 句法复杂度展示权力语法, 二者均反应知识编码的不同侧面, 可以作为学科知识水平和学术写作水平测量的有效方法。

关键词: 语义密度; 句法复杂度; 学科知识编码; 学科文化

[中图分类号] H319 [文献标识码] A [文章编号] 1674-8921-(2022)06-0055-13
[doi 编码] 10.3969/j.issn.1674-8921.2022.06.008

1. 引言

语义密度 (semantic density) 是澳大利亚悉尼大学社会语言学家 Karl Maton (2009, 2013) 提出的合法化语码理论 (Legitimation Code Theory) 中语义维度的一个语码模态。语义密度指社会文化实践中意义在符号 (包括术语、概念、短语、表达、体态、服饰等) 中的浓缩程度。与之相对的是语义引力 (semantic gravity), 指意义对语境的依赖。自从语义维度提出以来, 其与知识建构之间的关系得到了学者们的广泛关注。澳大利亚教育语言学家们注意到, 长期以来教师擅长通过语义引力手段解包抽象和复杂知识体系, 但很少注意到如何培养学生掌握浓缩意义并建构知识的手段。Maton 和 Matruglio (2009) 以及 Martin 等 (2010) 观察了初中历史课堂教学中教师如何引导学生逐渐减弱语义引力、加强语义密度来促进学生的累积型知识建构。Macnaught 等 (2013) 调查了 11 年级生物课教师的培训项目, 观察教师如何通过合作创作语篇的形式帮助学生

作者简介: 高彦梅, 北京大学外国语学院研究员、长聘副教授、博士生导师。主要研究方向为系统功能语言学、语义学、学科话语研究。电子邮箱: gaoyan@pku.edu.cn; 周江平, 西华师范大学外国语学院讲师、北京大学外国语学院外国语言学及应用语言学研究所博士生。主要研究方向为系统功能语言学、语法隐喻、语料库语言学。

* 本文系 2020 年北京市社会科学基金规划项目一般项目“学科英语研究的语法隐喻视角” (编号 20YYB007) 的阶段性成果。感谢匿名评审专家的修改建议。

引用信息: 高彦梅、周江平. 2022. 语义密度、句法复杂度与学科知识编码[J]. 当代外语研究 (6): 55-67.

学习知识建构和意义浓缩的方式。国内相关研究调查了教师课堂教学和教材中的语义密度。张德禄和覃玖英(2016)分析了中国教师外语教学课堂话语中的语义引力和语义密度,展示了教师如何通过课堂话语建构促进学生知识结构形成。吴奇格和朱永生(2016)调查了大学英语教材中三种不同语类的权力三项特征,发现科普语篇的权力三项特征最为明显,社会语篇的语义密度稍低,而文学语篇的权力三项特征不明显。该研究从教材的角度向大学英语教师展示了不同学科知识的编码特点。目前国内外研究中有关处于不同知识建构等级的知者(knower)学术语篇的编码特征研究较少。

合法化语码的语义维度关注了意义的浓缩和降解,主要涉及权力词汇(power words),尚未关注权力语法(power grammar)和权力组织(power composition)。句法复杂度(syntactic complexity)是语料库语言学和二语习得领域近年来开发的系统(Hunt 1965; Ortega 2003, 2015; Lu 2010, 2011; Lu & Ai 2015; Lu *et al.* 2020),用来测量第二语言学习者的英语作文质量和写作能力发展。该系统中涉及的句法复杂度指数直接反应了权力语法。国内句法复杂度的相关研究关注了中高级(硕士)、高级英语学习者(博士)和本族语者学术写作中句法复杂度的异同(Yang *et al.* 2015; 陆小飞、许琪 2016; 雷蕾 2017)。雷蕾(2017)的研究发现,与母语作者比较,中国英语学习者的名词复杂度与本族语者相当,但语言产出长度更短,从属结构也更少。目前国内外句法复杂度研究尚未关注到本科生的学术论文写作。

以往有关学科文化的研究显示,“学科文化差异反映在语言形式的差异上;同样地,语言形式的差异可预示学科文化的差异”(Becher 1987: 261; 杨越森等 2019:509)。学科文化不仅可以从特定句式结构来观察,术语和名词化的使用以及句法复杂度也反应学科知识的编码特征。参考 Becher 等人(Becher 1987; Becher & Trowler 2001)的学科认识论分析框架,杨越森等(2019)将语言学归入软科学。那么作为软科学的语言学学科的实践者如何通过语义密度和句法复杂度来编码自己的研究发现?本研究采用量化分析手段,观察本科语言学课程学习者(中级)、语言学专业研究生(中高级)、中国二语专业学者(高级)与英语母语作者(本族语)在相同学科领域学术论文摘要中的语义密度和句法复杂度分布,观察这两个参数与专业知识水平之间的对应状况,期待为学科知识编码特征观察提供借鉴。

2. 语义密度与句法复杂度

2.1 语义密度

语义密度是合法化语码理论中语义性维度的一个语码模态。该模态包括两个语码模态:语义密度和语义引力。Maton (2014: A-37)用语义密度来指社会文化实践中意义的浓缩程度。语义密度可以按照强度连续统分为相对较强(SD⁺)或较弱(SD⁻)。语义在实践中的浓缩程度越高,语义密度越强;浓缩的

意义越少,语义密度越弱。语义引力指将复杂知识通过举例或阐述进行具体说明,常常借助具有较高语境依赖的词汇和语法手段实现,如指示词、代词、时态情态表达等。语义引力强(SG⁺)意味着意义对语境的依赖程度高;语义引力弱(SG⁻),意味着意义对语境的依赖较弱。语义密度通过三个手段观察:措辞中的语义密度、小句中的语义浓缩和序列中的语义浓缩。其中措辞中的语义密度指整个语篇中的语义密度,可以通过包含较高语义密度的术语、抽象名词和名词化的整体分布来观察。小句中的语义浓缩指小句中包含较高语义密度词语多少。序列指小句复合体中所包含的具有较高密度值词语的分布。增强语义密度的方式包括:(1)将一连串的意义压缩为专业术语(technicality),如 phagocytosis, atherosclerosis;(2)将过程意义浓缩为抽象名词或名词化,如 inflammation, fusion, connection。语义密度主要由专业术语和名词化来体现。专业术语的语义密度高,一方面与其本身体现的专业性和抽象性有关,同时也与该学科话题范围有关(Martin 2013; 吴奇格、朱永生 2016)。Maton 和 Doran (2017) 提出了词汇语义密度量表,将不同词汇的语义密度等级做了细致划分,包括两个大类:技术类和日常类。每个大类又分别包括两个小类:技术类包括合成术语(如脂多糖 lipo-poly-saccharide)和单一术语(如淋巴 lymph);日常类包括名词化和普通词。语义密度由高到低的顺序为技术类合成术语、技术类单一术语、日常类名词化和日常类普通词。

名词化是概念语法隐喻的主要手段之一。Martin (2013) 在“内嵌识读能力:作为意义的知识”一文中提出了“权力三项”(power trio)概念,用来解释话语意义的建构。包括权力词汇、权力语法和权力话语组织(power composition),其中的权力词汇指的就是语义密度较高的专业术语(technicality);权力语法主要通过语法隐喻(grammaral metaphor),尤其是概念语法隐喻实现。概念语法隐喻包括经验语法隐喻和逻辑语法隐喻两类(Martin 2013)。经验语法隐喻指的是用隐喻的方式表达经验意义,其典型方式是用名词化形式(如例 1b 中的 death 和 smoking)来表达通常由动词(如例 1a 中的 die 和 smoke)体现的经验过程,或者将通常由形容词体现的属性(attribute)、评价等转换为一个参与者(participant)(吴奇格、朱永生 2016:52)。逻辑语法隐喻指将一致式中由两个小句呈现的过程序列转由动词或介词实现为小句内部的成分间关系,包括将时间、因果、条件、让步等关系转化为动词。如例(2)b 将例(2)a 中两个呈并列关系的过程转换为两个名词短语,连接关系也由并列关系转换为由介词实现的因果关系(本文例句中相关内容由下加横线和加粗标出)。由于隐喻化的名词通常具有较高的抽象等级和概括性,因此本研究将名词化表达归于语义密度测量。另外,由于逻辑隐喻主要由动词化和小句间关系体现,我们将其归入句法复杂度部分考察。

例(1)a. More people smoke, so more people die of cancer of the lung.

b. **Lung cancer death** rates are clearly associated with **increased smoking**.

(Halliday 2016: 32)

例(2)a. Blood flows more voluminosly and capillaries widen.

b. There is **increased blood flow** to the infected area **due to VASO-DISLATION** (widening of capillaries). (Martin 2013: 28)

语篇的语义密度体现为术语和名词化表达的出现频次。语义密度测量可以参考 Ure(1971)的实词密度计算公式。Ure 的公式通过计算实词与总词数的百分比获得。语义密度的计算可以通过专业术语和名词化与语篇总词数的百分比获得。测量公式如下:

$$SD = \frac{\text{专业术语数} + \text{名词化}}{\text{总词数}} \times 100$$

2.2 句法复杂度

二语习得研究领域的句法复杂度指语言产出中语言形式的变化和复杂化程度(Hunt 1965;雷蕾 2017:1)。Lu (2011)以及 Bulté 和 Housen (2012)等提出的句法复杂度计算方法包括两个维度和 23 个参数。两个维度是句法结构和句法复杂度指数。句法结构包括 9 个参数:总词数、句子、动词短语、小句、T 单位(T-unit)、依存小句、复杂 T 单位、并列短语和复杂名词。句法复杂度指数包括 14 个参数:句子平均长度(MLS)、T 单位平均长度(MLT)、小句平均长度(MLC)、每一个句子的小句数(C/S)、每一个 T 单位中的动词短语数(VP/T)、每一个 T 单位中的小句数(C/T)、每一个小句中的依存小句数(DC/C)、每一个 T 单位中依存小句数(DC/T)、每一个句子中的 T 单位数(T/S)、复杂 T 单位比率(CT/T)、每一个 T 单位中并列短语数(CP/T)、每一个小句中并列短语数(CP/C)、每一个 T 单位中的复合名词数(CN/T)、每一个小句中的复合名词数(CN/C)。其中 MLS、MIT、MLC 三个参数直接与逻辑隐喻有关。

2.3 研究问题

为了测量中级、中高级、高级英语学习者和本族语者在语义密度和句法复杂度维度是否存在差异,我们将研究问题归纳如下:

(1) 中国中级、中高级和高级学习者在学术论文摘要中所使用的语义密度与本族语学者学术论文摘要中的语义密度是否存在显著差异?

(2) 中国中级、中高级和高级学习者在句法复杂度方面与本族语作者之间是否存在显著差异?

(3) 语义密度和句法复杂度分别解释知识编码的哪个侧面? 是否可以作为测量知识等级的可操作性工具?

3. 研究方法

3.1 语料选取

本研究收集 2018—2020 期间语言学领域四组不同等级作者的学术论文摘要。(1) 语言学课程本科生(EFL)学术论文摘要 30 篇。本科生作者刚刚开始

学习如何撰写语言学方向的学术论文,尝试以研究者身份参与知识的解读与建构。(2) 语言学专业硕士生(MA)论文摘要 30 篇。通过三年专业训练,语言学硕士研究生(MA)已经具备了独立开展科研的能力。他们的硕士论文摘要体现了硕士论文研究课题的原创性贡献。(3) 语言学领域中国学者(CR)的论文摘要 30 篇。该组摘要选自语言学领域多位具有国际影响的中国学者的国际会议英文摘要,代表了中国学者在该领域的前沿学术水平。(4) 以英语为母语的顶级学者(NR)的论文摘要 30 篇。该组摘要选自语言学领域国际知名学者国际会议的学术论文。该组学者很多都是相关领域学术方向的引领者,代表本领域的国际最高水平。由于四组摘要长度不同,因此我们对统计结果做了标准化处理。

3.2 语义密度标注和句法复杂度计算

语义密度标注按照术语的抽象等级和使用频率判断。标注者为本领域具有较高英语专业知识水平和英文表达能力的学者。专业术语确定标准参考 Maton 和 Doran(2017)的词汇语义密度量表和语言学领域各方向术语词典。名词化概念语法隐喻的确定标准参考 Halliday(1994, 2016)、Martin (2013)和李文、郭建辉(2020)的判定标准,重点关注动词和形容词的名词化表达。两位研究者首先独立标注少量语篇,然后根据发现的具体问题商定统一标准。为了避免个人取向的主观化,正式语义密度标注由两位研究者分别完成。对于有差异的标注最后商定确定。句法复杂度计算采用 Lu (2010, 2011)、Ai 和 Lu (2013)以及 Lu 和 Ai(2015)开发的自动计算程序 L2SCA 获得。在统计结果基础上借助 R 语言测算组间差异。

4. 研究结果与讨论

4.1 语义密度和句法复杂度分布

语义密度测量我们采用了三个数值:专业术语、名词化和语义密度。通过手工标注专业术语、名词化表达,我们计算出四组摘要中的语义密度。四组语义密度平均值比较见图 1。

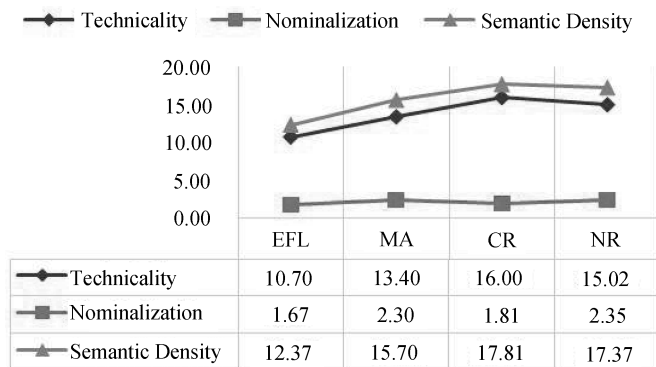


图 1 四组语义密度平均值比较

统计结果显示,语义密度均值随学术水平升高而呈微弱上升趋势:中国学者语义密度最高,但名词化使用与本科生持平;本族语学者语义密度第二,术语使用低于中国学者,名词化使用高于中国学者;硕士生语义密度第三,术语使用低于本族语学者和中国学者,名词化使用高于中国学者和本科生,与母语作者组持平;本科生组语义密度和术语使用最低,名词化使用与国内学者持平。采用 Mann-Whitey 检验分析结果显示:母语作者($p=0.001<0.01$)、中国学者($p=0.001<0.01$)和硕士生($p=0.001<0.01$)组的语义密度均显著高于本科生组;中国学者组显著高于硕士生组($p=0.037<0.05$);其余各组间差异不显著。四组语义密度差异见表 1。

表 1 四组语义密度差异

	MA	CR	NR
EFL	0.001**	0.001**	0.001**
MA		0.037*	0.097
CR			0.665

* $p<0.05$; ** $p<0.01$

句法复杂度数值主要参考句法复杂度参数,重点关注句子平均长度(MLS)、T 单位平均长度(MLT)、小句平均长度(MLC)等。句法复杂度计算结果在四组之间呈现梯级上升趋势:本科生组最低;硕士生组高于本科生组;中国学者组高于两个学生组;本族语组数值最高。四组数据趋势分布见图 2。

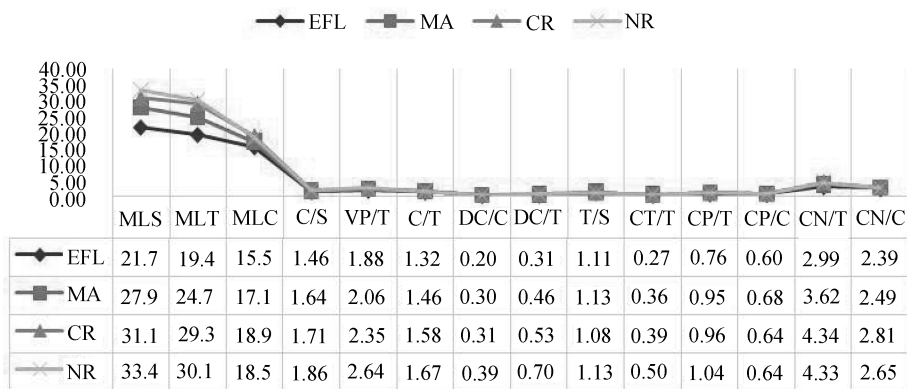


图 2 四组摘要句法复杂度平均值分布

同时我们也对这四组语篇句法复杂度的主要参数(包括 MLS, MLT 和 MLC)做了组间差异分析。由于四组语篇中大部分参数都呈非正态分布(除了硕士生(MA)组与母语学者(NR)组的 T 单位平均长度(MLT)参数外,前者 $p=0.082$,后者 $p=0.134$),我们采用了 Mann-Whitey 检验作为差异分析方法(结果见表 2)。表 2 显示:(1)从句子平均长度 MLS 和 T 单位平均长度 MLT 指标

看,本族语作者显著高于硕士生和本科生;中国学者和硕士生显著高于本科生;中国学者与本族语学者之间($p=0.337>0.05$; $p=0.701>0.05$)差异不显著;(2)从小句平均长度 MLC 指标看,中国学者显著高于本科生($p=0.033<0.05$),而其他组间差异不显著;(3)就语言产出长度的三个指标来看,中国学者与本族语学者之间差异不显著。

表 2 四组句法复杂度差异

	EFL-MA	EFL-CR	EFL-NR	MA-CR	MA-NR	CR-NR
MLS	0.020 *	0.000**	0.000**	0.170	0.021 *	0.337
MLT	0.015 *	0.000**	0.000**	0.039 *	0.015 *	0.701
MLC	0.314	0.033 *	0.063	0.253	0.390	0.775

* $p<0.05$; ** $p<0.01$

4.2 讨论

语义密度体现权力词汇和语法隐喻能力,句法复杂度与权力语法密切相关,二者从不同侧面体现知识编码特征。

(1) 语义密度显示权力词汇和语法隐喻能力

中国中级、中高级和高级学习者在学术论文摘要中所使用的语义密度与本族语学者学术论文中的语义密度存在显著差异。中国学者组的语义密度显著高于硕士生组和本科生组。本族语者与中国学者和硕士生组之间差异不显著。三个高水平组的语义密度显著高于本科生组。学术论文摘要展示一篇论文的主要贡献,体现该学科的文化特征。语言学领域的中国学者、本族语者和硕士生组在摘要中通过术语和名词化表达建构了自己的分析模型,提出自己的新见解、展示新知识的进展状况。三组较高水平作者的语义密度首先体现在技术类合成术语、单一术语和名词化行业词和通用词的使用。Martin (2013)、吴奇格和朱永生(2016)的研究表明,技术性和专业性较强的专业术语在语篇中通过构建类属关系帮助构建知识体系等级。术语之间的类属关系在中国学者和研究生产者的摘要中出现频率较高。

专业术语。学术摘要的一个重要功能就是展示自己的原创性成果,体现为对已有问题给出新的解决方案,对以往尚未研究的问题提出自己的分析框架,等等。在展示自己的新方案、新框架过程中,作者通常会运用语义密度较高的专业术语首先标示出自己研究的领域归属,然后构建出自己的研究框架。以中国学者组语义密度最高的 CR4 的片段为例,见例(3),我们可以发现,语义密度值与作者研究领域的学科术语密切相关。

例(3)

CR4: SFL provides an ergative model for the analysis of transitivity in English as a complement to the transitive model; for a generically ergative language like Classical Tibetan, the ergative model can be employed to interpret transitivity functions of a

clause. This paper is intended to offer a **systemic functional account** of the **ergative pattern** in **Classical Tibetan**, especially the various functions of the **ergative particle**. The major function of the **ergative case particle** is to identify the **Agent** in a clause if there is one, giving it some features of **circumstances** as the particle is also used to indicate **Cause** or **Manner**.^①

语篇开头由技术合成词形成的领域名称和模型名称具有较高的语义密度值,后面单一词构成的技术术语相对语义密度稍低,但高于名词化和细腻词。整个语篇的语义轮廓呈现由高向低的趋势。

名词化。除了专业术语,名词化也可以帮助建构作者研究的原创性。例(4)来自研究生组 MA4,作者介绍自己的最新成果基于新方法如何解决文本多标签分类中缺少高级别信息的问题。

例(4)

MA4: To tackle the problem of lacking in high-level information in **multi-label text classification**, **Semantic-Unit-based Dilated Convolution** constructs **dilated convolution** for the **extraction of semantic units** and a hybrid attention mechanism for the **fusion** of both semantic-unit-level information and word-level information. Experimental results show that it outperforms the baseline by +1.4 micro-F1 (%) and -0.09 **Hamming Loss** (%) in RCV1-v2 and +0.9 micro-F1 (%) and -0.11 **Hamming Loss** (%) in Ren-CECps.

作者首先通过复合技术名词 Semantic-Unit-based Dilated Convolution 和 hybrid attention mechanism 建构自己的复合模型,同时通过名词化概念语法隐喻介绍模型的工作原理 extraction of semantic units 和 fusion of both semantic-unit-level information and word-level information。从语义轮廓(semantic profile)来看,作者在段落开头、中间和结尾均呈现了较高语义密度。

与三组较高水平作者比较,语言学课程本科生的学术论文摘要的语义密度明显偏低。例(5)来自本科生组的 EFL5,在这篇摘要中作者尝试从历时角度观察法语对英语词汇的影响。

例(5)

EFL5: Many **French vocabulary** entered English after the **Norman Conquest** of England in 1066, when French became the language of the new **Anglo-Norman court**, the government, and of the elites for several centuries until after the **Hundred Years War**. From then on, English has continuously been impacted by the **French language**. A big part of the influence of **French** upon the English language pertains to its **lexicon**. More than a third of **modern English vocabulary** comes from French. In different periods, there are subtle differences in the ways of **borrowing**.

与上面的中国学者和硕士生比较,作者在提出两个具有较高语义密度的单一术语 French vocabulary 和 Norman Conquest 后,并没有按照读者预期那样继续展开重要时间节点,并解释法语对英语的影响。除概念语法隐喻 borrowing 之外,作者未能使用其他技术类和名词化手段。整个语篇的语义轮

廓呈现较低水平。

参考 Maton 和 Doran (2017), 上述三位作者的语义密度可以呈现图 3 的语义轮廓图。三个语篇都是在开始阶段呈现较高的语义密度, 在结尾阶段语义密度下降。

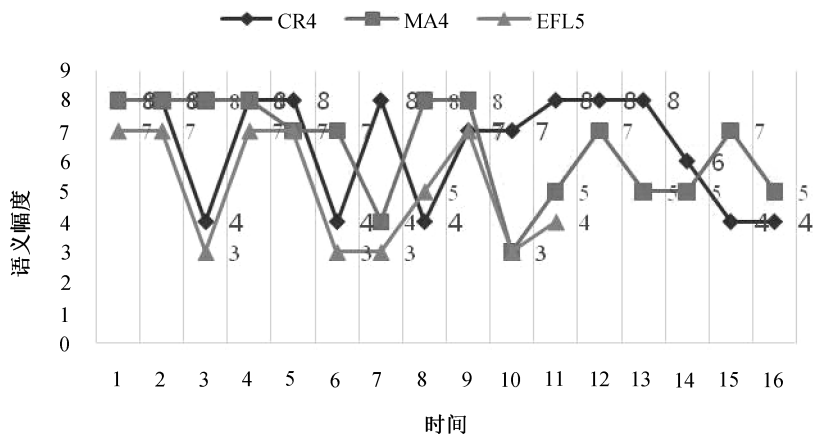


图 3 例(3)、(4)、(5)的语义轮廓图

(2) 句法复杂度显示权力语法和知识编码复杂度

学科文化特征不仅体现在语义密度, 同时也会体现在该学科的权力语法。本研究的四组语义密度分布显示中国学者和研究生在语义密度方面与母语作者没有显著差异。但权力语法的另一个参数即句法复杂度却呈现了另一幅图景: 母语作者组句法复杂度的所有指标, 尤其是前两个(句子平均长度(MLS)、T 单位平均长度(MLT))都显著高于其他三组。该结论与雷蕾(2017)和郭彩琛(2019)的发现一致, 即母语作者的语言产出长度指标均显著高于中国作者。雷蕾对比分析了中国硕士和博士论文与母语作者硕士和博士论文在句法复杂度的差异, 发现: “中国英语学习者写作中的从属结构和名词复杂度与本族语者相当, 但语言产出长度更短”(雷蕾 2017: 6)。以母语作者组中三个参数最高的 NR8 中的片段为例, 可以看到, 句子的平均长度与知识编码的复杂度关系密切。在例(6)中, 小句动词由下加横线和粗体标出。

例(6)

NR8: As evidence of the opportunities that are **opening up** for systemic functional approaches —for example through our emphasis on language variation, language evolution, and stylistic analysis —I **focus** on more than a decade of work which we **have conducted** at Macquarie University on the human responses to trauma, specifically those responses treated under the description of Borderline Personality Disorder (BPD). This **is** a particularly debilitating syndrome of dysregulation of emotions and distorted visions of oneself and the world (including the traumatic effects of, for example, schizophrenia, childhood abuse, anorexia; Haliburn 2017).

NR8 是母语作者组句法复杂度指标最高的语篇。例(6)为摘要中的一个独立段落,共 92 个词,只有两个句子,包含了 5 个动词短语、4 个小句、3 个独立单位(T-unit),2 个从属小句、3 个并列短语(Coordinate phrase)和 13 个复合名词短语(Complex nominal)。句法复杂度的三个指数分别为:平均句子长度 MLS=45.5,独立单位平均长度 MLT=30.3333,小句平均长度 MLC=22.75。我们将这个片段与 MA 组中一个类似长度的片段做比较,见例(7)。

例(7)

MA6: Thus, we **put forth** an umbrella term case variation and **use** it as a heading under which we **investigate** variant case marking patterns across languages. Of course, case variation **includes** phenomena of a much wider scope than the four mentioned above and we **will focus** on two-participant clauses in current study. It **is** the goal of this paper to offer a cognitive-functionally based framework with elaborations on case variation of two-participant clauses from a typological perspective.

该段落总字数为 76,包括 3 个句子、6 个动词短语、5 个小句、1 个并列短语和 13 个复合名词短语。在句法复杂度参数中,平均句子长度 MLS 为 25.3333,独立单位平均长度 MLT 为 19,小句平均长度 MLC 为 19,明显低于例(6)中的三个句法复杂度指标。

Ortega (2003,2015)、Lu(2010,2011)、Lu 和 Ai (2015)、雷蕾(2017)等国内外相关研究广泛证明,句法复杂度反应学术写作水平。Ortega (2003,2015)、Yang 等(2015)和 Kyle(2016)都发现,在写作水平与 T 单位平均长度(MLTU)之间具有显著正相关关系:随着写作者水平逐渐提高,T 单位平均长度呈上升趋势。本研究的结论可以对以往研究提出补充:平均句子长度和小句平均长度也反映学术写作水平和知识编码复杂度。

5. 结语

学科知识编码与很多因素有关,包括语言能力、科研能力、学术论文写作水平、学科知识水平等等。本文通过探讨语义密度、句法复杂度的分布,总结四组观察对象的学科知识编码特征发现以下三点:第一,语义密度和句法复杂度与知识编码水平呈现相同趋势,中高级和高级学习者与本族语作者的两组数据显著高于中级学习者;第二,中国高级学者语义密度显著高于中高级学者和中级作者,中高级学者显著高于中级作者,但高级学者与本族语学者以及中级学者与本族语学者之间差异不显著;第三,从语言产出长度三个指标观察,中级作者句法复杂度显著低于中高级学者、高级学者和本族语学者,高级学者与本族语学者之间差异不显著。调查结果表明,语义密度和句法复杂度从不同侧面反映知识编码特征,可以作为知识编码水平测量和学术写作能力测量的有效工具。

本研究对外语教育中的学科知识建构研究、对学术写作教学和基于句法复杂度的 L2 产出能力研究有多方面的启发。在语言学知识领域,不同等级知识

水平的学者对专业术语和名词化表达的驾驭不同,语义密度与知识水平呈现一致性。与本科生比较,中国学者、研究生和本族语作者更擅长通过学科领域技术类合成术语、单一术语和名词化来建构自己的学科归属,并可以通过不同类属关系展示自己研究的独特性。与此同时,句法复杂度与知识编码也具有 consistency 关系。中国学者与本族语作者在语言产出长度的三个指标上没有显著差异。研究生组句子平均长度和 T 单位平均长度显著低于中国学者和本族语作者。本科生的三个指标均显著低于研究生、中国学者和本族语作者。这一发现对学术写作有重要启发。在语言教学和学术写作过程中,教师可以有意识地引导学生分析学术论文中的语义密度分布以及复杂知识的编码特点,以促进学习者较早意识到不同学科知识的编码属性和特征。

Fadilah 和 Anugerahwati (2019)在总结语法隐喻发展历程基础上指出,高校学术阅读和写作教学中应注重培养学习者对专家作者语言风格的分析,应强化有关名词化、动词化表达的训练。本研究显示,本科生学术写作指导过程中也应注重引导学生关注学科领域的术语和概念语法隐喻(尤其是名词化)在知识建构中的关键作用。同时,本科和硕士培养过程中也要注意培养学生对语言产出长度的关注。学生可以学习如何通过词汇语法手段编码建构自己的知识结构,开拓和发展个人语义表达的资源库(repertoire)。

附注

① 此处为作者收集的语料,下文不再赘述。

参考文献

- Ai, H. & X. Lu. 2013. A corpus-based comparison of syntactic complexity in NNS and NS university students' writing [A]. In A. Díaz-Negrillo, N. Ballier & P. Thompson (eds.). *Automatic Treatment and Analysis of Learner Corpus Data* [C]. Amsterdam/Philadelphia: John Benjamins; 249-264.
- Becher, T. 1987. Disciplinary discourse [J]. *Studies in Higher Education* 12: 261-274.
- Becher, T. & P. Trowler. 2001. *Academic Tribes and Territories* [M]. Buckingham: Open University Press.
- Bulte, B. & A. Housen. 2012. Defining and operationalising L2 complexity [A]. In A. Housen, F. Kuiken & I. Vedder (eds.). *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA* [C]. Amsterdam: John Benjamins. 21-46.
- Fadilah, E. & M. Anugerahwati. 2019. Grammatical metaphor at tertiary level: Rise, development, and implications revisited [J]. *Argentinian Journal of Applied Linguistics* 7(1): 131-143.
- Halliday, M. A. K. 1994. *An Introduction to Functional Grammar* (2nd edn.) [M]. London: Edward Arnold.
- Halliday, M. A. K. 2016. *Aspects of Language and Learning* [M]. New York: Springer.
- Hunt, K. W. 1965. *Grammatical Structures Written at Three Grade Levels* [R]. NCTE

- Research Report No. 3. Champaign, IL: National Council of Teachers of English.
- Kyle, K. 2016. *Measuring Syntactic Development in L2 writing: Fine Grained Indices of Syntactic Complexity and Usage-based Indices of Syntactic Sophistication* (Unpublished Doctoral Dissertation) [D]. Atlanta: Georgia State University.
- Lu, X. 2010. Automatic analysis of syntactic complexity in second language writing [J]. *International Journal of Corpus Linguistics* 15(4): 474-496.
- Lu, X. 2011. A Corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development [J]. *TESOL Quarterly* 45(1): 36-62.
- Lu, X. & H. Ai. 2015. Syntactic Complexity in college-level English writing: Differences among writers with diverse L1 backgrounds [J]. *Journal of Second Language Writing* 29 (3): 16-27.
- Lu, X. J. E. Casal & Y. Liu. 2020. The rhetorical functions of syntactically complex sentences in social science research article introductions [J]. *Journal of English for Academic Purposes* 44: 1-16.
- Macnaught, L., K. Maton, J. R. Martin & E. Matruglio. 2013. Jointly constructing semantic waves: Implications for teacher training [J]. *Linguistics & Education* 24(1): 50-63.
- Martin, J. R. 2013. Embedded literacy: Knowledge as meaning [J]. *Linguistics and Education* 24: 23-37.
- Martin, J. R., K. Maton & E. Matruglio. 2010. Historical Cosmologies; Epistemology and Axiology in Australian Secondary School History Discourse [J]. *Revista Signos* 43 (74): 433-463.
- Maton, K. 2009. Cumulative and segmented learning: Exploring the role of curriculum structures in knowledge-building [J]. *British Journal of Sociology of Education* 30(1): 43-57.
- Maton, K. 2013. Making semantic waves: A key to cumulative knowledge-building [J]. *Linguistics and Education* 24(1): 8-22.
- Maton, K. 2014. A TALL order? Legitimation Code Theory for academic language and learning [J]. *Journal of Academic Language and Learning* 8(3): A34-A48.
- Maton, K. & Y. J. Doran. 2017. Semantic density: A translation device for revealing complexity of knowledge practices in discourse, part 1-wording [J]. *Onomázein* (1): 46-76.
- Maton, K. & E. Matruglio. 2009. How do we know: the social relation in school history discourse [R]. Sydney: The University of Sydney *LCT-SFL Roundtable*.
- Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing [J]. *Applied Linguistics* 24(4): 492-518.
- Ortega, L. 2015. Syntactic complexity in L2 writing: Progress and expansion [J]. *Journal of Second Language Writing* 29: 82-94.
- Ure, J. 1971. Lexical density and register differentiation [A]. In G. Perren & J. L. M. Trim (eds.). *Applications of Linguistics* [C]. London: Cambridge University Press. 443-452.
- Wolfe-Quintero, K., S. Inagaki & H. Y. Kim. 1998. *Second language development in writing: Measures of fluency, accuracy and Complexity* [M]. Hawaii: University of Hawaii Press.

- Yang, W., X. Lu & S. C. Weigle. 2015. Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality [J]. *Journal of Second Language Writing* 28: 53-67.
- 郭彩琛. 2019. 语言学会议论文摘要中的母语及二语作者语言特征研究[D]. 北京: 北京大学硕士学位论文.
- 雷蕾. 2017. 中国英语学习者学术写作句法复杂度研究[J]. 解放军外国语学院学报(5): 1-9.
- 李文、郭建辉. 2020. 中国高级英语学习者概念语法隐喻能力——基于中外博士论文的研究[J]. 外语教学理论与实践(1): 50-58.
- 陆小飞、许琪. 2016. 二语句法复杂度分析器及其在二语写作研究中的应用[J]. 外语教学与研究(3): 409-420.
- 吴奇格、朱永生. 2016. 大学英语教材中三个不同语类的“权力三项”特征分析[J]. 外语与外语教学(5): 50-58.
- 杨越森、苏杭、卫乃兴. 2019. 语义序列与学科文化探索[J]. 外语教学与研究(4): 508-520.
- 张德禄、覃玖英. 2016. 语义波理论及其在教师课堂话语分析和建构中的作用[J]. 外语教学(2): 52-55.
- 周惠、刘永兵. 2017. 英语学位论文摘要中语法隐喻的使用与语篇功能研究[J]. 现代外语(4): 484-494.

(责任编辑 邓梦寒)

Semantic Density, Syntactic Complexity and Disciplinary Knowledge Coding, by GAO Yanmei & ZHOU Jiangping.

Abstract: Semantic density and syntactic complexity reveal the characteristics of knowledge coding and culture of a discipline. The aim of this study is to testify the feasibility of semantic density and syntactic complexity as measures of knowledge coding in linguistics. By deploying the analytical tools of semantic density and syntactic complexity, this study compared four sets of abstracts of academic articles written by four groups of researchers at different academic levels, namely, experts of native speaker of English (NR), Chinese experts (CR), Chinese master degree students (MA), Chinese undergraduate students of linguistic courses (EFL). The results show that distributions of SD and SC are different among the four groups. NR, CR and MA used much more condensed technical terms than EFL writers. NR's syntactic complexity is much higher than that of the MA and EFL group. There is no significant difference between NR and CR. Data from the analysis testified the validity of the two analytical tools for assessing writers' capacity of knowledge coding.

Key Words: Semantic Density (SD); Syntactic Complexity (SC); disciplinary knowledge coding; disciplinary culture