

# 词类转化与语言形态类型关系的定量研究

邵斌 阎建玮 郑憬

**【摘要】**本文基于“劳特利奇系列频率词典”计算了8种语言各5000常用词中转类词的数量及比例,比较不同类型语言词类转化能产性的异同,同时基于滑动平均聚合度数值比较语言的形态丰富度,以此计算词类转化能产性与形态丰富度的相关性,从而考察词类转化与语言形态类型之间的关系。研究发现:1)转类能产性与语言形态类型密切相关,孤立语的转类能产性最高,强屈折语的最低,其他语言处于两者之间;2)基于滑动平均聚合度的计算表明,孤立语与屈折语在形态丰富度上构成连续统,语言形态类型具有梯度性,并非离散范畴;3)转类能产性与形态丰富度之间存在负相关性,即语言形态越贫乏则转类能产性越高,形态越丰富则转类能产性越低,此种负相关性反映了语言的词法与句法复杂度的折中关系,并证实了“复杂度折中假设”。

**【关键词】**词类;词类转化;语言形态类型;形态丰富度;复杂度折中假设

**【作者简介】**邵斌、阎建玮、郑憬,浙江大学外国语学院(杭州 310058)。

**【原文出处】**《外语教学与研究》(京),2023.4.497~508

**【基金项目】**本文为国家社科基金重点项目“基于语料库的汉英动名兼类词历时演变对比研究(1919-2019)”(20AYY001)的阶段性成果。

## 1. 引言

“词类转化”(conversion)简称转类,指的是词不改变形式便转化为另一种词类的构词现象,如英语的fish可由名词转为动词。词类转化在各种语言中普遍存在,Valera(2014)考察了64种语言,发现其中62种都存在转类现象。转类词通常需满足词形相同、词类不同和词义相关这三个条件。上述标准涉及词的形态、句法和语义层面,但三者常有扞格。因此,词类转化是语言研究的难点,Lieber(2005:418)指出,词类转化可视为各种语言学理论阵营多年来争论不休的战场。

传统的语言类型学根据形态对语言进行分类,通常把语言分成孤立语、屈折语、黏着语和多式综合语(陆丙甫、金立鑫2015:12)。对于转类与语言形态类型之间的关系,学界意见不一。有学者认为,缺乏形态变化的孤立语更易发生转类(Valera 2015:333),但也有学者指出,形态系统的减弱与词类转化的增加并无联系(Marchand 1960:295-296)。之所以存在上述分歧,原因有三:一是不同语言对转类的界定不尽相同,导致了跨语言研究的困难(Štekauer, Valera & Körtvélyessy 2012:219);二是缺乏转类能产性(productivity of conversion)的量化方法和数据,尚未有研究对转类能产性进行定量分析,也尚未有完善的测量方法(Valera 2015:333);三是前人在论述不同类型语言的转类现象时,未能对语言类型进行量化分析,而是视其为离散范畴,故结论不够精细。

有鉴于此,本文拟分别对转类能产性和语言形态类型进行计算,以此对两者的关系进行定量研究。具体而言,本文以8种语言各5000常用词中转类词的比例作为转类能产性指标,以基于大规模树库获得的形态丰富度作为语言形态类型的指标,通过计算两组数据间的相关性来探讨词类转化与语言形态类型之间的关系,

从而揭示不同类型语言词类转化的异同及规律。

## 2. 不同形态类型语言的词类转化界定及相关研究

本文所涉的语言分属孤立语和屈折语。孤立语是几乎没有屈折形态的语言,本文涉及汉语和越南语。屈折语有较为丰富的屈折形态,本文涉及德语、荷兰语、西班牙语、葡萄牙语和捷克语。按谱系,上述孤立语分属汉藏语系和南亚语系;屈折语同为印欧语系,但分属日耳曼、罗曼和斯拉夫语族。英语的形态类型较为复杂,古英语是典型的屈折语,但发展到现代英语,其屈折形态数量已经很少,因此有学者将其列为孤立语(Štekauer, Valera & Körtvélyessy 2012: 215; 陆丙甫、金立鑫 2015: 13)。但 Haselow(2011: 2)指出,英语并未成为一种完全的孤立语,因为孤立语无法通过黏着语素表达词类信息,而英语并非如此。Bauer, Lieber & Plag(2013: 616-617)也指出,英语的形态绝不能说是孤立的,很多词是形态复杂词,英语在类型上具有异质性(typologically heterogeneous)。由此可见,英语在形态类型上还不易归类。

转类现象在孤立语中很常见,不少学者甚至认为孤立语中词的兼类现象不宜称为词类转化,而是词类不定(categorical indeterminacy)。比如,Spencer(2005)指出,孤立语的词无需改变词形就可做名词或动词,此为词类不定,而非词类交叉。Bisang(2008)认为,上古汉语词无定类(pre-categoriality),即孤立的词并无名、动之分,只有入句才能显现其词类。然而,对于现代汉语而言,词无定类则有过度概括之嫌。吕叔湘、朱德熙(2013: 8)指出:“说汉语的词根本不能分类,这个话有点过甚其辞。汉语里的词能做两类或三类词用的究竟是少数,别种语言也有类似的情形。”现代汉语的词有较为明确的词类,这是转类的基础。从数据来看,胡明扬(1995)对 3036 个动词进行考察后发现,动名兼类词占动词总数的比例不到 20%。越南语中也有许多转类词,Nguyễn(1997: 83, 168-170)指出,Grammont 和 Lê Quang Trinh 等语法学家曾提出的“越南语没有词类”的观点与事实不符,事实是越南语与汉语类似,转类也只是少部分词的特征。

屈折语中也普遍存在转类现象,但屈折语对转类的界定与孤立语有所不同。Valera(2015)指出,屈折语的转类现象可分为两大类:一是狭义转类,这是严格的同形转类,如西班牙语形名转类词 *derecho*(正当的/权利);二是广义转类,可包括形式略有不同的异形转类。屈折语具有丰富的屈折形态,其名词或动词原形就可能带有屈折词缀。比如,德语的名词 *Antwort*(回答)和动词 *antworten*(回答)、荷兰语的名词 *hamer*(锤子)和动词 *hameren*(捶打)、西班牙语的名词 *aceite*(油)和动词 *aceitar*(涂油),上述动词后缀 *-en* 和 *-ar* 并非派生词缀,而是动词原形的屈折词缀。这种词对的词形虽稍有不同,但其词干一致,因此有学者认为屈折语的转类应视作词干层面发生的构词行为,上述例子便是基于词干的转类(stem-based conversion)(Štekauer, Valera & Körtvélyessy 2012; Lieber 2022: 58-60)。由上可知,孤立语的词类转化是以词为基础的(word-based conversion),属同形转类;屈折语的词类转化则包含同形转类和异形转类两种。

上文提及英语的类型归属存在争议,但英语中转类词数量众多也是事实(Jespersen 1924: 60-61; Crystal 1997: 295)。Pinker(1995: 379)甚至认为:“几个世纪以来,能把名词轻易转化为动词一直是英语语法的一部分,这是英语成其为英语的过程之一。”反过来,动词转成名词的现象在英语中也很常见。根据英语构词法传统,其转类构词都是以词为基础的同形转类,这是英语不同于典型屈折语的地方。然而,迄今为止,对词类转化能产性与语言类型关系的研究尚局限于定性分析。一般认为,孤立语的词类转化具有高能产性。在欧洲语言中,日耳曼语的词类转化能产性较高,比如英语以及稍弱一些的德语,而罗曼语和斯拉夫语的能产性较低(Valera 2015: 333)。但上述观点并非没有争议。Valera(同上: 330)指出,如果考察广义转类,那么英语的转类能产性就未必比西班牙语或拉丁语等强屈折语更高,屈折形态的消失不一定会导致转类的增加。之前有研究表

明,作为强屈折语的拉丁语、梵语和古英语都存在名动转类现象。简言之,转类能产性与语言形态类型之间的关系之所以不明晰,一是因为之前未有研究对转类能产性进行计算,故无法进行跨语言的定量比较,二是之前对语言类型定量分析的研究成果(如Greenberg 1960)未能用于考察转类现象。

有鉴于此,本文试图通过定量研究回答这一悬而未决的问题。本文的核心问题是词类转化与语言形态类型之间的关系如何。词类转化在此指其能产性程度,语言的形态类型由其形态丰富度所决定,因此该核心问题可转化为转类能产性与形态丰富度的相关性问题。转类能产性可通过词表中转类词的比例加以计算,形态丰富度可通过滑动平均聚合度(moving-average mean size of paradigm,简称MAMSP)加以计算,下文第3和第4节将对此详加讨论。

### 3. 跨语言的词类转化能产性

本文将词类转化能产性界定为常用词词表中转类词的数量及比例。这一界定的科学性在于:首先,这些词表均基于大型平衡语料库制作而成,例如,英语词表的制作以“当代美国英语语料库”(COCA)为基础;其次,各语言的词表具有可比性,它们都源自“劳特利奇系列频率词典”(Routledge Frequency Dictionaries),该系列词典收录基于语料库的5000常用词,可覆盖95%的书面语文本,反映语言使用的常态(Nation 1990);最后,以5000词中转类词的比例计算转类能产性是一种直观且可操作的科学方法。转类词的比例反映了该语言转类的难易程度,符合形态能产性的界定(Bolozky 1999:1)。该系列词典中,词形相同但词类不同的转类词按照频率分别列示。例如,英语词表中“动词cut”位列425,“名词cut”位列1493,cut即计为1个转类词型(type),2个转类词例(token)。需指出的是,“劳特利奇系列频率词典”不含越南语,本文的越南语词表源自以布朗语料库为参照的百万词级“越南语语料库”<sup>①</sup>。此外,劳特利奇汉语词表(标为“汉语1”)的语料库采用ICTCLAS工具进行自动分词和词类标注(Xiao, Rayson & McEnery 2009:3)。鉴于汉语分词和词类标注常有争议,本文还选取了“《人民日报》语料库”词表(标为“汉语2”)。《人民日报》语料库比较重视词的兼类标注,其所涉的兼类有七类,如名一量等,此外,还有vn、vd、an、ad四个标记,分别代表名动词(动词的名词用法)、副动词(动词的副词用法)、名形容词(形容词的名词用法)、副形容词(形容词的副词用法)。如“群众的支持”中的“支持”就标为名动词vn(俞士汶等2002)。虽然学界尚未就此处“支持”的词类达成共识,但汉语的“支持”“调查”等词显然与英语的support、survey等动名转类词类似,故本文将其视为转类词。

上述词表的转类词数量及占比见下页表1,其中基于词的转类词型(第5列)即同形转类词,如英语work、西班牙语deber、捷克语ryst等;基于词干的转类词型(第6列)是异形转类词,如荷兰语的名动词对besluit/besluiten,捷克语的名动词对boj/bojovat,-en和-ovat为其动词原形的屈折词缀;最后一列是基于词和词干的转类词型数量及比例之和。

由表1可知,第一,转类能产性与语言形态类型密切相关,孤立语与强屈折语构成转类能产性连续统。具体而言,孤立语的转类词比例最高,其转类词比例一般超过10%,英语的转类词比例与孤立语接近;其次是作为屈折语的日耳曼语(德语和荷兰语)和罗曼语(西班牙语和葡萄牙语),其转类词占比为4%~5%,再次是斯拉夫语的代表捷克语,其转类词比例为2.42%。进一步考察其差异显著性可发现,孤立语与屈折语的整体性差异具有统计显著性( $\chi^2=266.77, p<0.001$ )<sup>②</sup>。在屈折语内部,捷克语与其他4种语言的转类词比例具有显著差异( $\chi^2=25.83, p<0.001$ )。然而,日耳曼语族的德语、荷兰语与罗曼语族的西班牙语、葡萄牙语并无显著差异( $\chi^2=0.048, p>0.05$ )。前人定性研究认为,日耳曼语的转类构词能产性较高,罗曼语和斯拉夫语的能产性偏低(Valera 2015:333),而本文的定量分析则提供了较准确的细颗粒度描写:英语的转类能产性很高,而德语和荷

表1. 各语言转类词数量及占比

语言	总词例数	总词型数	单类词型数 及占比	转类词型数及 占比(基于词)	转类词型数及 占比(基于词干)	转类词型数及 占比(基于词+词干)
汉语1	5004	4640	4312(92.93%)	328(7.07%)		328(7.07%)
汉语2	5000	4502	3818(84.81%)	684(15.19%)		684(15.19%)
越南语	5078	4472	3954(88.41%)	518(11.59%)		518(11.59%)
英语	5005	4376	3785(86.49%)	591(13.51%)		591(13.51%)
德语	5078	4854	4632(95.43%)	68(1.40%)	154(3.17%)	222(4.57%)
荷兰语	5165	4959	4757(95.93%)	66(1.33%)	136(2.74%)	202(4.07%)
西班牙语	5008	4790	4572(95.45%)	52(1.08%)	166(3.47%)	218(4.55%)
葡萄牙语	5023	4817	4612(95.74%)	35(0.73%)	170(3.53%)	205(4.26%)
捷克语	5006	4908	4789(97.58%)	26(0.53%)	93(1.89%)	119(2.42%)

兰语等日耳曼语与西班牙语、葡萄牙语等罗曼语无明显差异,捷克语等斯拉夫语的能产性最低。第二,综合汉语和越南语的数据可知,孤立语中的转类词比例在10%~20%之间。“汉语1”的兼类词比例较低,原因可能是词类标注存在偏差,其5000词例中动名兼类词型数仅为84,而“汉语2”中的动名兼类词型数则为293,在汉语2中标为兼类词的“发展”“改革”“教育”“管理”“合作”等在汉语1中都只标为动词。故下文分析主要以汉语2为参照。

英语在转类能产性上表现出了与其他欧洲语言的差异。由表1可知,英语的转类能产性更接近于汉语和越南语。事实上,古英语演变为现代英语的过程中,其类型有着从屈折语向孤立语发展的趋势,词类转化从以词干为基础转向以词为基础(Kastovsky 1994)。从数据来看,英语的转类与汉语可谓大同小异,与其近邻语言德语和荷兰语却有较大差异。可见,在词类问题上,印欧语系语言的词类与汉语并非完全处于对立面,不存在“印欧语”与汉语的简单对立。需说明的是,不同语言的屈折形态差异只是数量上多与少的区别,并非有与无的绝对区别。事实上,孤立语和屈折语形成了一个连续统,语言形态类型具有原型性,并不是绝对离散的范畴。

#### 4. 跨语言的形态丰富度

孤立语与屈折语的本质区别就在于其形态系统是否丰富。因此,形态丰富度可作为两种类型语言的分类指标。Greenberg(1960)提出用基于每个词词素数量(morpheme/word ratio)的定量方法区分语言类型。然而,Wilson & Harvey(2020)认为,由于不同语言对词和词素的界定及其形态特征的外化方式存在差异,Greenberg的测量方法也不易操作。因此,本研究基于标注树库(Abeillé 2003),使用滑动平均聚合度(Yan & Liu 2021)计算上述语言的形态丰富度。MAMSP通过单位词元(lemma)中词形(form)的数量来定义形态丰富度(如mache、machst、macht、machen、machte、machend、mach、machtest、machtet等是同一德语词元MACHEN的不同词形)。MAMSP计算背后蕴含着语言类型学原理:理想的孤立语没有屈折变化,其词形/词元比为1。屈折语形态丰富,一个词元通常对应多个词形,其词形/词元比大于1。语言形态越丰富,词形/词元比就越大。MAMSP数值是考察语言类型的一个关键指标,该数值指标把语言置于“孤立语—屈折语”连续统中,考察其屈折程度,从而更客观地揭示该语言所属的类型:MAMSP数值越低,该语言便是越典型的孤立语;反之,则是越典型的屈折语。MAMSP的具体计算公式(同上:137)如下:

$$\text{MAMSP}(W) = \frac{\sum_{i=1}^{N-W+1} \frac{F_i}{L_i}}{N-W+1}$$

在该公式中, N代表树库中的字符数, W代表滑动窗口大小, N-W+1为滑动窗口数, F<sub>i</sub>是每个窗口中的词形数量, L<sub>i</sub>是词元数量,  $\frac{F_i}{L_i}$ 为词形/词元比,  $\sum_{i=1}^{N-w+1}$ 代表各窗口中词形/词元比之和, 即平均聚合度(MSP)之和。本文根据 Covington & McFall(2010), 以500个字符为标准滑动窗口计算 MAMSP。所依据的树库是通用依存标注数据集2.5版本(UD 2.5)(Zeman *et al.* 2019)。上述语言所涉23个树库<sup>③</sup>的MAMSP数值见表2。由表2可知, 整体而言, 孤立语的形态丰富度小于屈折语。孤立语中, 汉语和越南语的MAMSP值接近于1, 可见这两种语言的词形与词元基本一致。英语还保留了一定程度的屈折变化, 其MAMSP值为1.134, 与德语、荷兰语以及西班牙语、葡萄牙语接近。捷克语的数值为1.242, 表现出强屈折性。MAMSP数值体现了各语言在形态丰富度方面的差异, 也大体符合语言学家对上述各语言形态特征的定性判断。

表2. 各语言基于树库的滑动平均聚合度

语言	树库	语言编码 <sup>④</sup>	字符数	句子数	MAMSP	MAMSP均值
汉语	Chinese-CFL	zh_1	6,339	451	1.002	1.001
	Chinese-GSD	zh_2	106,226	4,997	1.001	
	Chinese-GSDSimp	zh_3	106,203	4,997	1.001	
越南语	Vietnamese-VTB	vi_1	37,431	3,000	1.000	1.000
英语	English-EWT	en_1	22,4964	16,622	1.153	1.134
	English-GUM	en_2	88,128	5,427	1.149	
	English-LinES	en_3	82,712	5,243	1.160	
	English-ParTUT	en_4	43,837	2,090	1.114	
	English-PUD	en_5	18,725	1,000	1.096	
德语	German-GSD	de_1	254,275	15,590	1.123	1.146
	German-HDT	de_2	3,003,124	189,928	1.121	
	German-LIT	de_3	35,034	1,922	1.192	
	German-PUD	de_4	18,558	1,000	1.148	
荷兰语	Dutch-Alpino	nl_1	185,853	13,578	1.107	1.100
	Dutch-LassySmall	nl_2	83,370	7,338	1.093	
西班牙语	Spanish-AnCora	es_1	484,060	17,680	1.146	1.133
	Spanish-GSD	es_2	384,062	16,013	1.119	
葡萄牙语	Portuguese-Bosque	pt_1	197,766	9,365	1.117	1.117
捷克语	Czech-CAC	cs_1	434,256	24,709	1.258	1.242
	Czech-CLTT	cs_2	31,106	1,125	1.310	
	Czech-FicTree	cs_3	135,261	12,760	1.257	
	Czech-PDT	cs_4	1,285,509	87,913	1.215	
	Czech-PUD	cs_5	15,986	1,000	1.169	

值得注意的是, 英语的形态丰富度明显大于作为典型孤立语的汉语和越南语, 其原因是现代英语仍保留了古英语的部分屈折变化。Lyovin, Kessler & Leben(2017: 22)指出, 英语既有孤立语的特征, 诸多词语无需经过词形变化便可名可动; 又有屈折语的特点, 如不规则动词通过元音变换体现屈折变化。因此, 英语在语言类型上是一个另类(typological anomaly)。一方面, 英语在转类能产性上接近于孤立语, 但另一方面, 英语在形态类型上则更接近屈折语。上述结果反映了英语具有部分孤立语和部分屈折语的特点, 介于孤立语与屈折语连续统之间。至此, 本文分别计算了转类能产性和形态丰富度, 下文将计算并讨论两者的相关性, 以此考察词

类转化与语言类型之间的关系。

### 5. 转类能产性与形态丰富度之间的相关性

由表1和表2可知,转类能产性和形态丰富度都不是离散变量,而是连续变量。基于表1和表2中的数据,本节考察转类能产性与形态丰富度这两组连续变量的相关性。研究使用斯皮尔曼秩相关系数 $\rho$ (Spearman's rank correlation efficient)衡量其相关性。如果 $\rho$ 为正值,则表示两者之间为正相关; $\rho$ 为负值,则为负相关。 $\rho$ 的绝对值越接近1,相关程度越高:若 $0.7 \leq |\rho| \leq 1$ ,相关程度高;若 $0.3 \leq |\rho| < 0.7$ ,相关程度中等;若 $0 \leq |\rho| < 0.3$ ,相关程度低(Levshina 2015: 119, 132)。图1a呈现基于词的转类词型数比例(即表1的第5列数据)与滑动平均聚合度(即表2的最后一列数据)的相关性,图1b为基于词和词干的转类词型数比例(即表1的第7列数据)与滑动平均聚合度的相关性。

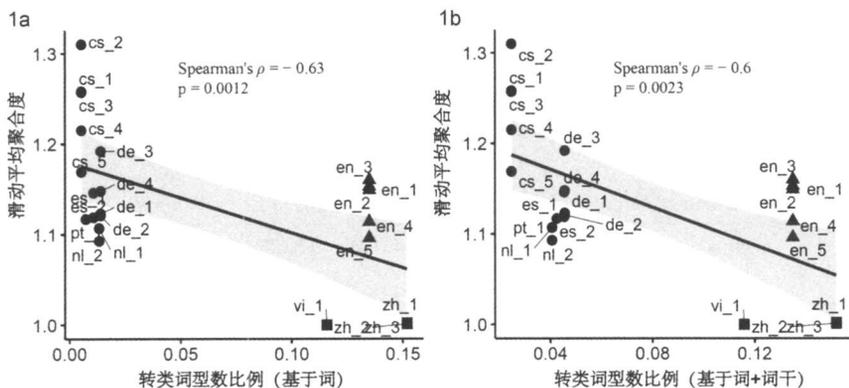


图1. 转类词型数比例与滑动平均聚合度的相关性

图1a显示,基于词的转类词型数比例与滑动平均聚合度之间存在中等负相关性( $\rho = -0.63, p < 0.01$ )。图1b显示,基于词和词干的转类词型数比例与滑动平均聚合度之间也存在中等负相关性( $\rho = -0.60, p < 0.01$ )。此外,由图1可知,典型的屈折语都聚类在左上角,典型的孤立语则聚类在右下角,说明转类词型数比例与滑动平均聚合度之间存在负相关性,也即转类能产性与形态丰富度之间存在负相关性。图1也呈现出英语在语言类型上的特殊性。其位置既不在典型的屈折语所处的左上角,也不在典型的孤立语所处的右下角,而是处于右中部位置。事实上,如果暂不考虑英语,其他7种语言的狭义和广义转类能产性与形态丰富度之间便呈现出高度的负相关性, $\rho$ 分别为 $-0.77(p < 0.001)$ 和 $-0.73(p < 0.001)$ 。换言之,其他语言在语言类型上表现出了更强的典型性。

对于词类转化能产性与形态丰富度之间的负相关性,也有学者持质疑态度。比如, Marchand(1960: 295-296)认为,屈折系统的减弱与词类转化的增加并无联系。拉丁语是高度屈折的语言,但拥有大量的名词转动词。然而,以上论断是定性判断,并非定量研究的结果。本文则用定量数据证明,词类转化能产性与形态丰富度之间确实存在负相关性。

事实上,语序与转类能产性及形态丰富度密切相关。孤立语在很大程度上依赖语序和语境而非形态特征来表达词在句中的语法关系。汉语的形态贫乏,但其语序较为固定,即便两个词的词形一致,但因其在中位置不同,其功能也容易区分。换言之,汉语即便存在大量的转类词,但其较为固定的语序保证了句子信息的准确传达。英语虽非典型的孤立语,但在该问题上也与汉语类似。英语史学者也指出,中古英语的屈折形态逐渐脱落,名词的主语或宾语功能无法借助屈折形态加以区分,只能诉诸语序,因此SVO语序就成为英语的主导

语序,该语序能较好地区分名词的主语和宾语或施事和受事的功能(Barber, Beal & Shaw 2009: 170)。因为语序传达了词语之间的语法关系,即便同形的转类词也不妨碍对句子信息的正确解读。比如,汉语句子“小偷偷偷东西”连续出现了4个“偷”,英语句子 Love loves to love love. 中出现了4个 love,但SVO语序排除了其歧义。语言类型学家考察了848种语言发现,屈折形态贫乏的语言更倾向于使用SVO语序(Sinnemäki 2010),其原因便在于此。而屈折语往往依赖屈折形态来表达词在句中的语法功能。比如,斯拉夫语的形态已经表明了词与词的关系,语序在句中所起的作用不如孤立语那么重要,因此具有较高的自由度。汉语和英语的语序固定,比如“狗咬人”和 dog bites man 两句中的主语和宾语不能互换位置,否则其义截然相反。但捷克语的语序则非常灵活,捷克语的 člověk 意为“人”,当其做动词的直接宾语时变位为 člověka, -a 即是宾格标记。捷克语中“狗咬人”既可以用SVO语序,即 Pes kouše člověka,也可以用OVS语序,即 Člověka kouše pes,其意义并无改变。可见,形态丰富度、转类能产性与语序自由度三者之间存在一种动态关系:形态丰富的语言可借助形态而非语序来体现词与词之间的语法关系,故其语序自由度高,词类转化能产性低。反之,形态贫乏的语言语序自由度低,词类转化能产性高。

上述结果证实了语言中的“复杂度折中假设”(complexity trade-off hypothesis),即“负相关性假设”(negative correlation hypothesis)。该假设认为人类语言的相关要素在复杂度方面存在负相关性(Crystal 1997: 6; Coloma 2017)。Hockett(1958: 180-181)指出,任何语言的语法(包含词法和句法)的整体复杂性应当大致相同,因为所有语言都有大体同样复杂的内容要表达,因此不能通过词法表达的内容就得通过句法来表达。换言之,如果词法复杂,则句法简单;反之,如果句法复杂,则词法简单。Crystal(1997: 5, 6)也指出,所有语言都有一个复杂的语法体系:它可能在某一方面相对简单(如无词尾),但通常它在另一方面又会相对复杂(如语序),即语言是一个自动调节的均衡系统。Koplenig *et al.*(2017)基于大规模语言数据调查发现,通过语序表达的信息量与通过词的内部结构所表达的信息量符合折中假设:对语序信息依赖程度越高的语言对词内结构信息的依赖程度就越低。换言之,如果词内结构携带的信息较少,则语序必须提供更多的信息才能成功地进行交流。形态丰富度与词类转化能产性的负相关性,其实质就是词法复杂度与句法复杂度的折中关系。

转类能产性与形态丰富度的负相关性也表明语言是一个复杂适应系统(Köhler 2005)。在这个复杂适应系统中,各语言要素之间是一种自组织、自适应的动态平衡关系,词法和句法是这个复杂系统中互相协同的重要参项。本研究发现,孤立语倾向使用更为固化的语序来表示语法关系,因此语言使用者需要消耗更多的认知负荷去编码语序信息。这就意味着他们无法超负荷去编码复杂的形态特征,最终导致孤立语的形态编码会相对简单。相反,屈折语的语序限制并不严格,语言使用者的认知负荷更多用于编码词的形态特征。正如 Koplenig *et al.*(2017: 4)所述,形态与语序之间的这种动态平衡是“省力原则”的表现(Zipf 1949)。就语言复杂适应系统而言,本文的实证结果为人类如何编码语言信息以进行有效交流提供了新的参考。

综上所述,本研究揭示了不同语言对词类和形态信息的编码方式,证明了语言的句法与词法复杂度之间存在折中关系,即人类语言中相关参项的复杂度之间呈负相关性,也验证了人类语言中普遍存在的“省力原则”。

## 6. 结论

本文计算了8种语言的词类转化能产性与形态丰富度之间的相关性,以此考察词类转化与语言形态类型的关系。转类能产性与语言形态类型密切相关,孤立语与强屈折语构成能产性梯度,孤立语的转类能产性最高,强屈折语的最低,其他语言介于两者之间;孤立语与屈折语在形态丰富度上构成连续统,语言形态类型具

有梯度性,并非离散范畴。总体而言,转类能产性与形态丰富度之间存在负相关性,这与语序自由度有关,语言的形态贫乏,则语序较为固定,转类能产性就较高,此种负相关性反映了语言的词法与句法复杂度的折中关系,并证实了“复杂度折中假设”。这种折中关系的背后是语言作为复杂适应系统其内部要素之间的动态平衡关系,因此本研究为自然语言如何编码信息以实现有效沟通提供了新的参考。

在判别各种语言的转类词时,请教了以下老师和同学:Magda Ševčíková、Hana Hledíková、练斐、左雅、史慧琳、沈莹、曾凡珂、彭雁飞,特此致谢。感谢《外语教学与研究》编辑部及匿名审稿专家提出的宝贵修改意见。

#### 注释:

- ①本越南语语料库及频率词表由红河学院刘克强教授研制并准许本研究使用,特此致谢。
- ②此处整体性差异考察的是2种典型孤立语转类词型平均数与5种屈折语转类词型平均数之间的差异。
- ③UD 2.5中涉及该8种语言的树库共24个,其中英语的English-Pronouns树库中的句子为语法书中的例句,并非真实使用状态下的语言,故本文未将其收录在内。
- ④8种语言的编码取自ISO 639-1标准,详见[https://en.wikipedia.org/wiki/List\\_of\\_ISO\\_639-1\\_codes](https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes)。

#### 参考文献:

- [1]Abeillé, A. 2003. *Treebanks: Building and Using Parsed Corpora*[M]. Dordrecht: Kluwer Academic Publishers.
- [2]Barber, C., J. Beal & P. Shaw. 2009. *The English Language: A Historical Introduction* (2nd edition)[M]. Cambridge: CUP.
- [3]Bauer, L., R. Lieber & I. Plag. 2013. *The Oxford Reference Guide to English Morphology*[M]. Oxford: OUP.
- [4]Bisang, W. 2008. Precategoriality and syntax-based parts of speech: The case of Late Archaic Chinese[J]. *Studies in Language* 32 (3): 568-589.
- [5]Boložky, S. 1999. *Measuring Productivity in Word Formation*[M]. Leiden: Brill.
- [6]Coloma, G. 2017. The existence of negative correlation between linguistic measures across languages[J]. *Corpus Linguistics and Linguistic Theory* 13(1): 1-26.
- [7]Covington, M. & J. McFall. 2010. Cutting the Gordian knot: The moving-average type-token ratio(MATTR) [J]. *Journal of Quantitative Linguistics* 17(2): 94-100.
- [8]Crystal, D.(ed.). 1997. *The Cambridge Encyclopedia of Language* (2nd edition)[Z]. Cambridge: CUP.
- [9]Greenberg, J. 1960. A quantitative approach to the morphological typology of language[J]. *International Journal of American Linguistics* 26(3): 178-194.
- [10]Haselow, A. 2011. *Typological Changes in the Lexicon. Analytic Tendencies in English Noun Formation*[M]. Berlin: Walter de Gruyter.
- [11]Hockett, C. 1958. *A Course in Modern Linguistics*[M]. New York: The Macmillan Company.
- [12]Hu, Mingyang[胡明扬].1995.A quantitative study on Chinese verb-noun heterosemy[J]. *Studies in Language and Linguistics*(2): 91-99.[动名兼类的计量考察,《语言研究》2]
- [13]Jespersen, O. 1924. *The Philosophy of Grammar*[M]. London: George Allen & Unwin.
- [14]Kastovsky, D. 1994. Typological differences between English and German morphology and their causes[A]. In T. Swan, E. Mørck & O. Westvik(eds.). *Language Change and Language Structure: Older Germanic Languages in a Comparative Perspective*[C]. Berlin: Mouton de Gruyter. 135-157.

- [15]Koplenig, A., et al. 2017. The statistical trade-off between word order and word structure—Large-scale evidence for the principle of least effort[J]. PLoS ONE 12(3): Article No. e0173614.
- [16]Köhler, R. 2005. Synergetic linguistics[A]. In R. Köhler, G. Altmann & R. Piotrowski(eds.). Quantitative Linguistics: An International Handbook[C]. Berlin: Mouton de Gruyter. 760–774.
- [17]Levshina, N. 2015. How to Do Linguistics with R[M]. Amsterdam: John Benjamins.
- [18]Lieber, R. 2005. English word formation processes[A]. In P. Štekauer & R. Lieber(eds.). Handbook of Word-formation[C]. Dordrecht: Springer. 375–422.
- [19]Lieber, R. 2022. Introducing Morphology[M]. Cambridge: CUP.
- [20]Lu, Bingfu & Lixin Jin[陆丙甫、金立鑫].2015. An Introduction to Linguistic Typology[M]. Beijing: Peking University Press.[《语言类型学教程》。北京:北京大学出版社]
- [21]Lü, Shuxiang & Dexi Zhu[吕叔湘、朱德熙].2013. Talks on Grammar and Rhetoric[M]. Beijing: The Commercial Press.[《语法修辞讲话》。北京:商务印书馆]
- [22]Lyovin, A., B. Kessler & W. Leben. 2017. An Introduction to the Languages of the World[M]. Oxford: OUP.
- [23]Marchand, H. 1960. The Categories and Types of Present-Day English Word-Formation: A Synchronic-Diachronic Approach[M]. Wiesbaden: Otto Harrassowitz.
- [24]Nation, I. 1990. Teaching and Learning Vocabulary[M]. New York: Newbury House.
- [25]Nguyễn, Đình-Hoà. 1997. Vietnamese[M]. Amsterdam: John Benjamins.
- [26]Pinker, S. 1995. The Language Instinct: How the Mind Creates Language[M]. New York: Harper Perennial.
- [27]Sinnemäki, K. 2010. Word order in zero-marking languages[J]. Studies in Language 34(4): 869–912.
- [28]Spencer, A. 2005. Towards a typology of ‘mixed categories’[A]. In C. Orgun & P. Sells(eds.). Morphology and the Web of Grammar [C]. Stanford: CSLI Publications. 95–138.
- [29]Štekauer, P., S. Valera & L. Körtvélyessy. 2012. Word-formation in the Worlds Languages: A Typological Survey[M]. Cambridge: CUP.
- [30]Valera, S. 2014. Conversion[A]. In R. Lieber & P. Štekauer(eds.). The Oxford Handbook of Derivational Morphology[C]. Oxford: OUP. 154–168.
- [31]Valera, S. 2015. Conversion[A]. In P. Müller et al.(eds.). Word-formation: An International Handbook of the Languages of Europe. Vol. 1[C]. Berlin: De Gruyter Mouton. 322–339.
- [32]Wilson, A. & R. Harvey. 2020. Using rank-frequency and type-token statistics to compare morphological typology in the Celtic languages[J]. Journal of Quantitative Linguistics 27(2): 159–186.
- [33]Xiao, Richard, P. Rayson & T. McEnery(eds.). 2009. A Frequency Dictionary of Mandarin Chinese: Core Vocabulary for Learners [Z]. London: Routledge.
- [34]Yan, Jianwei & Haitao Liu. 2021. Morphology and word order in Slavic languages: Insights from annotated corpora[J]. Voprosy Jazykoznanija 70(4): 131–159.
- [35]Yu, Shiwen, et al.[俞士汶等]. 2002. The basic processing of contemporary Chinese corpus at Peking University SPECIFICATION [J]. Journal of Chinese Information Processing(5): 49–64.[北京大学现代汉语语料库基本加工规范,《中文信息学报》5]
- [36]Zeman, D., et al. 2019. Universal Dependencies 2.5[OL]. <http://hdl.handle.net/11234/1-3105>(accessed 20/12/2019).
- [37]Zipf, G. 1949. Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology[M]. Cambridge, MA.: Addison-Wesley.